Stochastic Optimization for Learning-based Super-resolution: Algorithms and Applications

Jun Zheng

Department of Computer Science

APPROVED:

______________________________
Olac Fuentes, Chair, Ph.D.

______________________________
Luc Longpré, Ph.D.

______________________________
Ming-Ying Leung, Ph.D.

______________________________
Abhijeet Ghosh, Ph.D.

______________________________
Patricia D. Witherspoon, Ph.D.
Dean of the Graduate School
©Copyright

by

Jun Zheng

2010
to my

FAMILY and FRIENDS

for their love, support, and sacrifices
Stochastic Optimization for Learning-based Super-resolution: Algorithms and Applications

by

Jun Zheng

THESIS
Presented to the Faculty of the Graduate School of
The University of Texas at El Paso
in Partial Fulfillment
of the Requirements
for the Degree of
Doctor of Philosophy

Department of Computer Science
THE UNIVERSITY OF TEXAS AT EL PASO
December 2010
Acknowledgements

More than twenty years journey of study through three schools, two universities, and two countries has led me to this point. I have experienced joy, pain, doubt, excitement, stress, struggle and so on. Fortunately I have teachers, friends, advisors, and families to help, support, and guide me along this journey. Without them I would never be here.

First I would like to thank my advisor, Dr. Olac Fuentes of the Computer Science Department at The University of Texas at El Paso (UTEP), for his continuous support, guidance, understanding, encouragement, and enduring patience during my doctoral study and research. The teaching he gave me will benefit me forever. I am sincerely grateful for the opportunity to study with him.

I also wish to thank my committee, Dr. Luc Longpré of the Computer Science Department, Dr. Ming-Ying Leung of the Mathematical Sciences Department and the Bioinformatics Program, both at UTEP, and Dr. Abhijeet Ghosh of Computer Science Department at University of Southern California. Their suggestions, comments and additional guidance were invaluable to the completion of this work. Furthermore, I want to thank all of the members of Vision and Learning Lab (VL^2). I have enjoyed working with such an energetic group of people.

Additionally, I want to thank professors and staff of UTEP Computer Science Department for all their teaching, hard work and dedication, providing me excellent environment to complete my study.

And finally, I must thank my family for their never-ending love and faith in me throughout the years. Most importantly of all, I want to thank my dear wife for her encouraging and supporting me to pursue my dream.
Abstract

Human beings get much of their information visually and depend on perception of images for many critical tasks, such as object identification, medical image analysis, photography, etc. In many visual-based applications, higher resolution images are required for perceiving and receiving critical information. A high resolution image can contribute to a better identification of a suspect’s face, or a more accurate localization of a tumor in a mammogram, or a more pleasing view in high definition television, and so on. However, it is hard to obtain the high resolution images needed for some applications, for example, the cost of sensors increases as a exponential function of their resolution so that a high resolution sensor may be too expensive. Another example is in X-ray imaging for medicine. Higher incident X-ray beam intensity produces higher resolution images but harms patients. In some cases image resolution can be improved through image super-resolution, an image processing procedure that takes a degraded image or a sequence of images as input and produces an image or a sequence of images of higher quality as output.

This dissertation addresses the problem of optimization of super-resolution algorithm according to the specific requirements of the applications for which the images are used.

One application addressed in this thesis is super-resolution for surveillance videos. In surveillance applications, cameras are usually set up with wide fields of view to capture as much of the scene as possible. This normally results in low-resolution images of suspects’ faces, therefore face image super-resolution is critical for face recognition tasks. This thesis improves the efficiency of face image super-resolution using stochastic search for local modeling by exploiting the fact that face patches maintain relatively tight distributions for shape at successive iterations. Furthermore, though face super-resolution could improve the appearance of face images dramatically, the detailed facial features such as eyes, eyebrows, nose, mouth and teeth of the super-resolution face are different from the ground truth. This thesis studies whether face super-resolution can help face identification by either human beings or computers, and proposes a novel patch-based simultaneous face identification and super-resolution approach that integrates
face identification and super-resolution together.

Another application addressed in this thesis is super-resolution for medical imaging. The high-quality mammogram is the most effective technology presently available for breast cancer screening. This thesis studies mammogram super-resolution, or synthesizing a high-resolution mammogram from an input low-resolution mammogram. Additionally, as medical imaging moves towards complete digital imaging and produces very large amounts of data, compression is necessary for storage and communication purposes. Super-resolution has the potential to eliminate the need to store images at full resolution, since they could be re-generated from low-resolution ones. This thesis explores mammogram compression using super-resolution, and then proposes a novel hybrid compression method for mammograms using super-resolution algorithms as a post-processing of the compression process. To further determine whether compression affects clinical diagnostic performance, this thesis studies whether these differences would affect micro-calcification detection by applying a computer-aided micro-calcification detection system to the original images and compressed images respectively, and then comparing the detection rates.

Lastly, this thesis considers super-resolution for 3D face scanning. Detailed facial geometry contributes significantly to the visual realism of face models in computer games, movies, virtual reality applications, and so on. This thesis proposes a new technique for real-time high-resolution 3D facial scanning using stochastic super-resolution to generate a specular normal map based on the diffuse normal map, instead of capturing both of them during scanning process. This thesis also demonstrates that the technique can be also used to transfer specular normal information to unpolarized data, or transfer specular normal information from one person to other people, or the synthesized normal map can be used for image-based lighting.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>v</td>
</tr>
<tr>
<td>Abstract</td>
<td>vi</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>viii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td><strong>Chapter</strong></td>
<td></td>
</tr>
<tr>
<td>1  Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 The definition of resolution</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Image resizing</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Super-resolution</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Main contributions</td>
<td>6</td>
</tr>
<tr>
<td>1.5 Organization of this thesis</td>
<td>8</td>
</tr>
<tr>
<td>2  Literature Review</td>
<td>9</td>
</tr>
<tr>
<td>2.1 Super-resolution</td>
<td>9</td>
</tr>
<tr>
<td>2.1.1 Single-frame super-resolution</td>
<td>10</td>
</tr>
<tr>
<td>2.1.2 Multi-frame super-resolution</td>
<td>12</td>
</tr>
<tr>
<td>2.1.3 Learning-based super-resolution</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Super-resolution for medical imaging</td>
<td>14</td>
</tr>
<tr>
<td>2.3 Super-resolution for 3D face scanning</td>
<td>15</td>
</tr>
<tr>
<td>2.4 Chapter summary</td>
<td>17</td>
</tr>
<tr>
<td>3  A Stochastic Method for Face Image Super-Resolution</td>
<td>19</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>19</td>
</tr>
<tr>
<td>3.2 Framework</td>
<td>20</td>
</tr>
<tr>
<td>3.3 Global modeling</td>
<td>21</td>
</tr>
</tbody>
</table>
3.4 Local modeling ......................................................... 22
3.5 Stochastic search in local modeling ......................... 23
3.6 Experimental results ................................................. 25
3.7 Chapter summary ....................................................... 28

4 Patch-based Simultaneous Face Super-resolution and Recognition ............. 33
4.1 Introduction .......................................................... 33
4.2 Simultaneous face super-resolution and identification ....................... 34
4.3 Experiments .......................................................... 36
4.4 Chapter summary ....................................................... 38

5 Super-resolution of Mammograms ........................................ 43
5.1 Introduction .......................................................... 43
5.2 Related work .......................................................... 44
5.3 Super-resolution algorithm ........................................... 45
5.4 Experimental results ................................................... 47
5.5 Chapter summary ....................................................... 49

6 Hybrid Compression of Mammograms Using Super-resolution .................. 55
6.1 Introduction .......................................................... 55
6.2 Automatic segmentation ............................................... 56
   6.2.1 Global threshold segmentation ............................... 57
   6.2.2 Region-growing segmentation ............................... 58
6.3 Huffman coding-based JPEG near lossless compression ....................... 59
   6.3.1 JPEG compression ............................................... 60
   6.3.2 Huffman coding-based JPEG near lossless compression ........... 61
6.4 Compression using super-resolution .................................... 62
6.5 Evaluation ............................................................ 64
6.6 Experimental results .................................................... 66
6.7 Chapter summary ....................................................... 72
# List of Tables

3.1 Quantitative evaluation ......................................................... 27
3.2 Computational time ............................................................. 28

5.1 Image quality measures ........................................................ 50
5.2 Comparison of PSNR of our methods with other methods ............... 50

6.1 Average results of compression with super-resolution for 10 mammograms compared with JPEG 2000 and lossless JPEG ...................... 68
6.2 10-fold cross validation microcalcification detection results on MIAS dataset ...................................................... 72
6.3 Huffman coding-based JPEG near lossless compression results on MIAS dataset ...................................................... 72
6.4 Hybrid compression results on MIAS dataset ................................ 73
List of Figures

1.1 The IEEE standard STD 208-1995 “Measurement of Resolution of Camera Systems” target ................................................. 4

3.1 Faces in surveillance images .................................................. 19
3.2 Face hallucination framework ............................................... 21
3.3 Local modeling with stochastic search .................................... 23
3.4 Similar faces are connected in a graph structure ...................... 25
3.5 23 manually selected facial feature points ................................. 26
3.6 (a) Original high-resolution face images. (b) Input low-resolution face images. (c) Generated super-resolution face images. (d) Results of bicubic interpolation. (e) Global faces. (f) Local faces. This figure shows the super-resolution results of exhaustive search. Compared with the input low-resolution images and the bicubic interpolation results, the super-resolution face images have much clearer detailed features. ................................................................. 29
3.7 (a) Original high-resolution face images. (b) Input low-resolution face images. (c) Generated super-resolution face images. (d) Results of bicubic interpolation. (e) Global faces. (f) Local faces. This figure shows the results of super-resolution using stochastic search. The size of randomly generated search set is set to 300. However, this is too small, so that the quality of local face decreases. .................. 30
3.8 (a) Original high-resolution face images. (b) Input low-resolution face images. (c) Generated super-resolution face images. (d) Results of bicubic interpolation. (e) Global faces. (f) Local faces. In this figure, the size of randomly generated search set is set to 1200. ......................................................... 30
3.9 (a) Original high-resolution face images. (b) Input low-resolution face images. (c) Generated super-resolution face images. (d) Results of bicubic interpolation. (e) Global faces. (f) Local faces. In this figure the size of randomly generated search set is set to 500.

3.10 (a) Original high-resolution face images. (b) Input low-resolution face images. (c) Generated super-resolution face images. (d) Results of bicubic interpolation. (e) Global faces. (f) Local faces. In this figure, the size of randomly generated search set is set to 2000. From Figure 3.7 to 3.10, we can see that stochastic local modeling with a larger generated search set would have better quality of local faces but take a longer time. Therefore, a tradeoff between quality and efficiency need to be found.

4.1 Low-resolution face images

4.2 Sample faces for training from AT&T dataset

4.3 Super-resolution results (a) Super-resolution face. (b) Global face. (c) Input low-resolution face image. (d) Original high-resolution face image.

4.4 Super-resolution results (a) Super-resolution face. (b) Global face. (c) Input low-resolution face image. (d) Original high-resolution face image.

4.5 Super-resolution results (a) Super-resolution face. (b) Global face. (c) Input low-resolution face image. (d) Original high-resolution face image.

4.6 Face identification results of 10-fold cross validation on AT&T dataset

4.7 Face identification results of low-resolution images, interpolated images and super-resolution images on AT&T dataset using neural networks

5.1 Automatic registration: (a) Original image. (b) Converting the original image to binary image. (c) Using erosion to remove the labels in the binary image. (d) Converting the binary image to gradient image to find the edge. (e) Using skeletonization to reduce the edge to a single line. (f) Sampling points on the line and making the mesh.
5.2 Sample super-resolution results with a downsample factor of 2: (a) Original high-resolution images. (b) Super-resolution images. (c) Interpolated images. (d) low-resolution images. ................................................................. 51
5.3 Sample super-resolution results with a downsample factor of 4: (a) Original high-resolution images. (b) Super-resolution images. (c) Interpolated images. (d) low-resolution images. ................................................................. 52
5.4 Sample super-resolution results with a downsample factor of 8: (a) Original high-resolution images. (b) Super-resolution images. (c) Interpolated images. (d) low-resolution images. ................................................................. 53
5.5 Sample super-resolution results with a downsample factor of 16: (a) Original high-resolution images. (b) Super-resolution images. (c) Interpolated images. (d) low-resolution images. ................................................................. 54
6.1 Automatic segmentation and compression of mammograms .................. 57
6.2 Global Threshold segmentation: (a) Original mammogram (b) Segmented mammogram after applying global threshold. (c) Segmented mammogram after removing label regions. ................................................................. 58
6.3 Region-growing Segmentation: (a) Input mammogram. (b) Binary gradient mask. (c) Dilated mask. (d) Filled mask. (e) Eroded mask. (f) Region of interest. .... 59
6.4 Flowchart of the method ................................................................. 63
6.5 Mammograms with labels of micro-calcification from the MIAS dataset, the window that contains the micro-calcification, and the corresponding window enhanced by Nonsubsampled Contourlet Transform .................................................. 67
6.6 Gray-Level Difference Method (GLDM) : Four Cumulative Distribution Functions (CDFs) for four different displacement vectors of the input image. ....... 68
6.7 PSNR of the decompressed images after super-resolution .................... 69
6.8 MSSIM of the decompressed images after super-resolution .................. 70
6.9 A subtraction of original image from the decompressed image. (a) Original image. (b) The decompressed images after hybrid compression (Coeffs=15; downsample factor=16). (c) subtraction of original image from the decompressed image.

6.10 Sample results of compression with super-resolution of different resolutions. (a) Original image. (b) The decompressed images after super-resolution (Coeffs=1, 2, 3, 5, 10, 15; downsample factor=4). (c) The decompressed images after super-resolution (Coeffs=1, 2, 3, 5, 10, 15; downsample factor=8). (d) The decompressed images after super-resolution (Coeffs=1, 2, 3, 5, 10, 15; downsample factor=16).

6.11 Sample result of region-growing segmentation. The red area is the ROI which includes important clinical information, and the blue circle encloses micro-calcification.

6.12 Micro-calcification coverage results of automatic region-growing segmentation.

6.13 Sample results of hybrid compression. (a) Original image. (b) JPEG compressed image (The number of DCT coefficients used for compression from top to bottom is 1, 2, 4, 8, 16, 32, and 64 respectively). (c) Compressed image using hybrid compression (Coeffs=15, downsample factor=16).

6.14 A close observation of the sample results of hybrid compression. (a) Original image. (b) JPEG compressed image (The number of DCT coefficients used for compression from top to bottom is 1, 2, 4, 8, 16, 32, and 64 respectively). (c) Compressed image using hybrid compression (Coeffs=15, downsample factor=16).

7.1 Setup hardware for face scanning.

7.2 The projector projects a sequence of four colored strip patterns and one uniform white pattern.

7.3 The primary light apparatus we use to create the gradient lighting patterns consists of about 150 LED lights placed on the vertices and edges of a twice-divided icosahedron.

7.4 (a) The camera calibration process uses the camera to observe a planar shown at a few different orientations. (b) The planar for calibration.
7.5 Move the planar freely to show at a few different orientations.  

7.6 Projected structured light patterns. (Top) Images captured by the left camera. (Bottom) Images captured by the right camera. $P_k$ is the stereo image for colored strip pattern $k$, and $W$ is the stereo image for the uniform white pattern.  

7.7 Structured light ratio images. (Top) Images captured by the left camera. (Bottom) Images captured by the right camera.  

7.8 Captured spherical gradient illumination images. (Top) Images with specular reflections. (Middle) Images with diffuse reflections. (Bottom) The four spherical gradient lighting conditions reflected in a mirrored sphere. (a) A linear gradient along the x-coordinate. (b) A linear gradient along the y-coordinate. (c) A linear gradient along the z-coordinate. (d) A constant pattern.  

7.9 A linear polarizers is mounted on a servomotor in front the camera, which enables the polarizer to be rapidly flipped on its diagonal between horizontal and vertical orientations.  

7.10 Normal acquisition from spherical gradient illumination. (a) Diffuse normal map. (b) Specular normal map.  

7.11 A high dynamic range (HDR) face dataset which consists of floating-point images of about $1944 \times 1296$ pixels with a frontal view of about 15 different expressions.  

7.12 Manually drawn facial masks. (a) Face image. (b) Facial mask.  

7.13 (a) Diffuse normal map. (b) Synthesized specular normal map. (c) Captured specular normal map. This figure shows the specular normal map synthesis results of the same person with different expressions. The synthesized specular normal map has much clearer detailed features than diffuse normal map, and almost the same as the captured specular normal map.  

7.14 (a) Base geometry. (b) Geometry optimized with synthesized specular normal map. (c) Geometry optimized with captured specular normal map. This figure shows the geometry optimization results of the same person with different expressions.
7.15 Renderings of face scans with different illumination sources. (a) Grace Cathedral. (b) Uffizi Gallery. (c) Pisa courtyard. This figure shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map. The results show that our synthesized specular normal maps can also work with image-based lighting techniques.

7.16 (a) Diffuse normal map. (b) Synthesized specular normal map. (c) Captured specular normal map. This figure shows the specular normal map synthesis results of cross-object transfer. The algorithm transfers one person’s specular information to another person.

7.17 (a) Base geometry. (b) Geometry optimized with synthesized specular normal map. (c) Geometry optimized with captured specular normal map. This figure shows the geometry optimization results of cross-object transfer.

7.18 Renderings of face scans with different illumination sources. (a) Grace Cathedral. (b) Uffizi Gallery. (c) Pisa courtyard. This figure shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map which is transferred from another person.

7.19 (a) Mixed normal map. (b) Synthesized specular normal map. (c) Captured specular normal map. This figure shows the specular normal map synthesis results of the same person with different expressions. The synthesized specular normal map are synthesized from unpolarized mixed normal map.

7.20 (a) Base geometry. (b) Geometry optimized with synthesized specular normal map. (c) Geometry optimized with captured mixed normal map. This figure shows the geometry optimization results of the same person with different expressions.
7.21 Renderings of face scans with different illumination sources. (a) Grace Cathedral. (b) Uffizi Gallery. (c) Pisa courtyard. This figure shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map. The results show that our synthesized specular normal maps can also work with image-based lighting techniques.
Chapter 1

Introduction

The majority of information received by human beings is visual, and most visual-based applications can benefit from high-resolution images. A high-resolution image can contribute to a better identification of a suspect’s face in a surveillance video, or to a more accurate localization of a tumor in a mammogram, or could produce a more pleasing view in high definition televisions, or generate a clearer image for industrial inspection, or make a better image for remote sensing and so on. However, sometimes it is very expensive to obtain high-resolution images, for example, the resolution of an image taken by a camera is dependent on the resolution of the CCD sensor, as the resolution of the image generated by a sensor increases, so does the cost of the sensor and therefore it may be too expensive. Another example is in X-ray imaging applied to medicine. To obtain a better quality image the incident X-ray beam intensity should be increased, but this would be harmful to the health of patients.

Image super-resolution is a set of image processing procedures that take a degraded image or a sequence of images as input and produce an image or a sequence of images of higher quality [IP90]. However, the definition of degraded images is subjective and mostly depends on the application. The degraded images can be low-resolution surveillance videos, or mammograms obtained with very low dosage of radiation, or even compressed images with very high compression ratio. The concept of image quality can also be defined in various ways. For example, in photography, quality refers to the appreciation of a human observer, while in object detection tasks quality may refer to good localization of edges. Based on the definition of degraded image and image quality, the super-resolution algorithms as well as the results may greatly vary.

The question addressed in this dissertation is whether super-resolution algorithms can be tailored to the specific requirements of the application for which the images are used. In this thesis
I develop super-resolution algorithms that can recover degraded images according to specific requirements of the application for which the images are used.

Super-Resolution (SR) has been applied to a variety of fields, such as video surveillance, medical imaging, high definition television, remote sensing, cell phone digital cameras and so on. In video surveillance systems, a location of interest is monitored by webcams or camcorders to detect abnormalities. The video is usually saved into video clips, and the system triggers various alerts such as sending an email, or setting off an alarm, or even enabling users to watch the video through Internet by video streaming. Because of the large amount of data and bandwidth constraints, the video frames are often captured at very low rate. And because cameras are usually set up with wide-fields of view to capture as much of the scene as possible, images of the objects of interest are usually in low-resolution. Therefore SR is needed both temporally and spatially for security purposes such as personal identification and authentication.

Medical imaging represents another area where SR has been successfully applied, in particular, in magnetic resonance imaging (MRI), mammography, position emission tomography (PET) and so on. For example, each year many women are diagnosed with breast cancer, which is one of the leading causes of cancer death for females worldwide. But if breast cancer can be detected early, the five-year survival rate would increase considerably. High-quality mammography, which uses X-rays to examine the human breast, is the most effective technology presently available for breast cancer screening. However, using high dosage radiation can increase the risk of patients while using low dosage radiation would lead to low-resolution images, which have low signal-to-noise ratios (SNR). Fortunately, the image resolution can be enhanced by super-resolution techniques.

SR techniques are also being applied to image and video compression. Compressed video is rapidly becoming the preferred method for video delivery, such as Internet streaming, wireless videophones, HDTV and so on. However, these applications require a significant amount of compression for commercial use, and the compression process is always implemented in a lossy manner which may end up producing images that look ”blocky”. This is acceptable with many applications, but when an application requires a high-resolution image, a super-resolution
algorithm must be employed.

1.1 The definition of resolution

Resolution can be measured in various ways. Researchers in digital image processing and computer vision use the term as follows:

**Pixel resolution** People often use the term pixel resolution as pixel count in a digital image, which is a vector of two positive integer numbers. The first number is the number of pixel columns, and the second is the number of pixel rows, for example 256 by 128. Alternatively, people describe pixel resolution as the total number of pixels in an image, typically given as number of megapixels, which is calculated by multiplying pixel columns by pixel rows and dividing by one million [Lyo06].

**Spatial resolution** Spatial resolution refers to the spacing of pixels in an unit length or area, expressed as dots per inch, pixels per inch, lines per millimeter, pixels per square inch and so on. It measures how closely pixels can be resolved in an image. The higher the spatial resolution, the greater the number of pixels in an unit area and correspondingly, the smaller the size of individual pixels. Figure 1.1 shows the IEEE standard STD 208-1995 “Measurement of Resolution of Camera Systems” target, which is a typical test target used to determine the spatial resolution of imaging sensors and imaging systems.

**Brightness resolution** Brightness resolution determines the number of brightness levels of a pixels. A pixel of gray images usually has 256 brightness levels which are represented by 8 bits. For RGB color images, each color channel has 8 bits, so that there are 24 bits in total for one pixel.

**Temporal resolution** Temporal resolution is also known as frame rate which refers to the numbers of frames captured per second in a movie or video. The typical frame rate suitable for a pleasing view is about 15 to 30 frames per second, while high-speed cameras may resolve 100 to 1000 frames per second, or even more.
In this dissertation, the term resolution unequivocally refers to the spatial resolution, and super-resolution is the process of obtaining a high-resolution image from a low-resolution input.

1.2 Image resizing

Increasing the number of pixels in an image or resizing of the image doesn’t necessarily increase the resolution of the image, because it doesn’t produce the high-frequency components lost during the down sampling process. People usually refer to the process of resizing as upsampling or image zooming. The typical method of upsampling is interpolation.

The simplest interpolation method is the nearest-neighbor algorithm which just replaces the interpolated point with the nearest neighboring pixel. The advantage is its simplicity but it tends to produce images with a blocky appearance.

Bilinear interpolation or bicubic interpolation can provide more satisfactory results. Bilinear
interpolation is a weighted average of four neighboring pixel values, which, mathematically, can be written as

\[ Q(x, y) = w_1 Q(x_1, y_1) + w_2 Q(x_1, y_2) + w_3 Q(x_2, y_1) + w_4 Q(x_2, y_2) \]

where \( Q(x, y) \) is the intensity value at the interpolated point \((x, y)\), and \( Q(x_1, y_1), Q(x_1, y_2), Q(x_2, y_1), \) and \( Q(x_2, y_2) \) are the intensity values of the 4 neighboring pixels.

The bicubic interpolation uses 16 neighboring pixels to estimate the interpolated point. It approximates the local intensity values using a bicubic polynomial surface. The bicubic interpolated image is smoother than corresponding images obtained by bilinear interpolation or nearest-neighbor interpolation but it is more computationally expensive than both of them. The general form for a bicubic interpolation is as follows:

\[ Q(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^i y^j \]

where \( Q(x, y) \) is the intensity value at the interpolated point \((x, y)\). The bicubic interpolation problem consists of determining the 16 coefficients \( a_{ij} \), as explained in [Key81].

### 1.3 Super-resolution

The amount of information available in the input image limits the quality of the image generated by interpolation algorithms. Because image interpolation cannot produce the high frequency components lost during the down-sampling process, researchers rarely consider interpolation as a super-resolution algorithm. To further improve the quality of low-resolution images,
reconstruction-based super-resolution algorithms require the investigation of multi-input data sets in which additional data constraints from several low-resolution images of the same scene can be used. Huang and Tsai [HT84] first proposed a frequency domain approach, and then numerous super-resolution algorithms in this line have been proposed [PKS87, KP88, KB90, IP91, TAR93, MP94, PST97]. However, nearly all these SR algorithms are based on the fundamental constraints that the super-resolution image should generate the low-resolution input images when appropriately warped and down-sampled to model the image formation process. In [BK02], Baker and Kanade show that the reconstruction constraints provide less and less useful information as the magnification factor increases, and for large enough magnification factors any smoothness prior leads to overly smooth results with very little high-frequency content, no matter how many low-resolution input images are used.

To break this limitation of reconstruction-based super-resolution, a new family of super-resolution approaches, based on statistical machine learning, aiming at analyzing large data sets of images of a particular class of objects and learning the mapping from low-quality to high-quality images of that class, have been proposed [LLX+01, LSZ01, WT03, LL04, LLT05, WT05, Ded07, LSF07] dating back to the hallucination approach of Baker and Kanade [BK00]. The learning-based algorithm can infer, for example, significantly better high-resolution face images depicting the same person as a low-resolution image given as input.

1.4 Main contributions

The driving motivation of the different problems addressed in this thesis is unified under the name of learning-based super-resolution. In this thesis I develop algorithms that can enhance the resolution of degraded images according to specific requirements of the application for which the images are used. The main contributions of this thesis are:

1. **Efficiency of face image super-resolution using stochastic search.** One problem addressed in this thesis is super-resolution of surveillance videos. I improve the efficiency of face image super-resolution using stochastic search for local modeling. Experimental
results show that the proposed algorithm generates high-quality face images from low-resolution inputs while reducing the computation time dramatically.

2. **Simultaneous face image super-resolution and recognition.** Though face super-resolution could improve the appearance of face images dramatically, the detailed facial features such as eyes, eyebrows, nose, mouth and teeth of the super-resolution face are different from the ground truth. In this thesis, I study whether face super-resolution can help face recognition by either computers or human beings, and propose a simultaneous face super-resolution and recognition algorithm. The experiments show that face super-resolution can not improve the performance of conventional face identification algorithms, but my method, which combines face super-resolution and face identification together, can help face identification by computers in low-resolution images.

3. **Super-resolution of mammograms.** Another problem addressed in this work is super-resolution of mammograms. I study mammogram super-resolution, which synthesizes a high-resolution mammogram from an low-resolution input automatically, with the help of a large collection of other high-resolution mammograms from many individuals, and study whether super-resolution can help automatic micro-calcification detection. The experimental results show that the super-resolution algorithm can generate high-quality high-resolution breast mammographies from low-resolution input with no manual registration.

4. **Hybrid compression of mammograms using super-resolution.** The problem of recovering a high-resolution image from an compressed image or a sequence of compressed images is also considered in this thesis, which describes a novel hybrid compression method for mammograms. I first use automatic segmentation to identify the clinically important areas. Then, a modified version of Huffman coding-based JPEG lossless compression is applied in such a way that the extracted areas from the first step are compressed losslessly, while the remaining regions are compressed lossily by downsampling before applying the encoding procedure, and applying super-resolution techniques after the decoding procedure to recover the original resolution image. Experimental results show that this method
provides a high compression ratio of about 62:1 while fully preserving about 90% of the marked micro-calcification. Additionally, the micro-calcification detection system can obtain almost the same detection rates on our decompressed mammograms as on original mammograms.

5. Specular Normal Synthesis Using Stochastic Super-resolution for Detailed Facial Geometry

This thesis proposes a new technique for real-time high-resolution 3D facial scanning using stochastic super-resolution to generate specular normal map based on diffuse normal map, instead of capturing both of them during scanning process. I analyze a training dataset of diffuse normal maps and specular normals of a particular object to learn the mapping from low-frequency components of diffuse normal maps to high-frequency components of specular normal maps of that object, and then infer the most likely high-resolution specular normal map detail depicting the same person as a low-resolution diffuse normal map given as input. Experimental results show that the proposed algorithm generates high-quality specular normal maps from diffuse normal map inputs.

1.5 Organization of this thesis

This dissertation is divided into nine chapters. In Chapter 1, I present an overview of the related work relevant to the area of super-resolution. Chapter 2 gives a high-level overview of the related work. Then I directly present my work in each subsequent chapter. Chapter 3 discusses super-resolution optimization for face images. And chapter 4 presents a simultaneous face super-resolution and recognition algorithm. Chapter 5 discusses super-resolution optimization for mammogram. Chapter 6 discusses a novel hybrid compression method for mammograms using super-resolution. Chapter 7 presents my work on specular normal synthesis using stochastic super-resolution for high-resolution 3D facial scanning. In chapter 8, I close this dissertation with a summary of my contributions and several directions for future work.
Chapter 2

Literature Review

2.1 Super-resolution

Existing methods for image super-resolution can be divided into two categories: reconstruction-based super-resolution and learning-based super-resolution. The problems under the reconstruction-based super-resolution category can be single-frame or multi-frame super-resolution. Single-frame super-resolution aims to estimate missing high-resolution details from a single input low-resolution image. However, they have resulted in blurring of sharp edges, introduction of blocking artifacts, inability to generate high frequency components or fine details of semantically important structures [Dod97, ABA01, GAA00, FPC00]. In multiple-frame super-resolution, the LR frames typically depict the same scene. That means that LR frames are distorted as well as shifted with subpixel precision. If the LR frames contain different subpixel shifts from each other, then the new information contained in each LR frame can be used to construct an HR frame. Through motion analysis from frame to frame, a super-resolution image can be inferred by combining these LR frames within subpixel accuracy [PPK03].

Reconstruction-based super-resolution provides less and less useful information as the magnification factor increases, and for large enough magnification factors any smoothness prior leads to overly smooth results with very little high-frequency content, no matter how many low-resolution input images are used [BK02]. To break this limitation of reconstruction-based super-resolution, a new family of super-resolution approaches, based on statistical machine learning, aiming at analyzing large data sets of images of a particular class of objects and learning the mapping from low-quality to high-quality images of that class, have been proposed. This enables them to infer significantly better high-resolution images as a low-resolution image given as input.
In this chapter, I briefly review the available literature for super-resolution algorithms.

### 2.1.1 Single-frame super-resolution

Earlier work of single-frame super-resolution addresses the problem of reconstructing a high-resolution image from a sequence of low resolution images [HT84]. However, it assumes a purely translational motion, and estimates the relative shifts between the observations and then infers the samples on a uniform grid with a higher sampling rate, and is based on low-resolution images without degradation and noise.

Ur and Gross [UG92] performed a non-uniform interpolation of a group of spatially shifted low-resolution images based on the generalized multichannel sampling theorem. Irani et al. [IP91] proposed an iterative algorithm together with a method for image registration with sub-pixel accuracy to increase image resolution. The approach is similar to reconstruction of a 2-D object from its 1-D projections in computer aided tomography (CAT). Images are reconstructed from their projections in many directions in tomography, while in super-resolution, each low-resolution pixel is a projection of a region in the scene whose size is determined by the imaging blur. The high-resolution image is then constructed using an approach similar to the back-projection method used in CAT.

To reconstruct a high-resolution noise-free image from multiple frames of down-sampled low-resolution noisy images, a weighted recursive last-squares theory based algorithm was developed by Kim et al. [KBV90], using the aliasing relationship between the down-sampled frames and the reference image. Later, Kim and Su [KS93] extended this method to provide a unified approach for deblurring, noise removal, and high-resolution reconstruction of low-resolution images with different kind of blurs and noises.

In [KIAS93, KAIS93], Komatsu et al. applied the non-uniform sampling theorem to transform non-uniformly spaced samples obtained by multiple cameras to a single uniform sampling grid. Originally, they used multiple cameras with the same pixel aperture, but there were severe
limitations both in the arrangement of multiple cameras and in the configuration of the scene, to
guarantee the spatial uniformity of the resultant resolution. To overcome these limitations, they
used multiple cameras with different pixel apertures, and developed a new, alternately iterative
signal processing algorithm that could be applied to different aperture cases.

Previous work has either solved registration, restoration, and interpolation independently, or
more recently, solved either the first two steps (registration and restoration) or the last two steps
together. However, none of the existing methods solve all three sub-problems simultaneously,
Tom and Katsaggelos [Tk95] posed the super-resolution problem as a maximum likelihood (ML)
problem which was solved by the expectation-maximization (EM) algorithm. By exploiting the
structure of the matrices involved, the problem was solved in the discrete frequency domain. The
ML problem was then the estimation of the sub-pixel shifts, the noise variances of each image,
the power spectra of the high resolution image, and the high resolution image itself.

Chiang and Boult [CB00] show that image warping has a strong impact on the quality of im-
age super-resolution. By coupling the degradation model of the imaging system directly into the
integrating resampler, they could better approximate the warping characteristics of real sensors,
which also significantly improved the quality of super-resolution images.

Stark and Oskoui [SO89] considered the problem of reconstructing remotely obtained images
from image-plane detector arrays, and suggested convex projections as an alternative to matrix
inversion. Although the individual detectors may be larger than the blur spot of the imaging
optics, high-resolution reconstructions can be obtained by scanning or rotating the image with
respect to the detector. They used convex projections to show that readily obtained prior knowl-
edge could be used to obtain good-quality imagery with reduced data. Tekalp et al. [TOS92]
extended this method to take into account the presence of both sensor blurring and observation
noise. They proposed a new two-step procedure, and illustrated that the projection onto convex
sets (POCS) formulation presented for the high-resolution image reconstruction problem could
also be used as a new method for the restoration of spatially invariant blurred images. Further
more, Calle and Montanvert [CM98] considered the problem of resolution enhancement as an
inverse problem of image reduction. The image to be enhanced has to be part of a set of images
that are similar to the image when being reduced. A projection of any image onto this set provides one of the possible enlarged images, which they called an induction of the image onto a set of acceptable super-resolutions. They proposed two methods of accurate projections which must be used according to the consistence of induced set.

2.1.2 Multi-frame super-resolution

Most multi-frame super-resolution algorithms are extended from single-frame super-resolution algorithms. Irani et al. [IP91] first proposed an iterative algorithm together with a method for image registration with subpixel accuracy to increase single image resolution. Later, they extended this method for image sequences by accurately computing image motion [IP93]. After computing the motion for different image regions, these regions are enhanced by fusing several successive frames covering the same region.

Schultz and Stevenson [SS96] proposed a new observation model based on motion compensated subsampling for a video sequence, and used Bayesian restoration with a discontinuity-preserving prior image model to extract a high-resolution video frame given a short low-resolution sequence, which is an extension to their earlier work on single image expansion [SS94].

Bascle et al. [BBZ96] proposed a method to track the object of interest through the sequence using region based matching, and then reconstructed a high-resolution deblurred image from a low-resolution sequence taking into account pixel sampling and motion blur.

Hong et al. [HKK97] introduced an iterative regularized approach to obtain a high resolution video sequence. This method defined a multiple input smoothing convex function which was used to obtain a globally optimal high resolution video sequence, and described a mathematical model of multiple inputs by using the point spread function between the original and bilinearly interpolated images in the spatial domain, and motion estimation between frames in the temporal domain.

Patti et al. [PST97] proposed a model which captures video with an arbitrary input sampling lattice and a non-zero aperture time. Based on this model they proposed an algorithm
using the theory of projections onto convex sets to reconstruct a super-resolution image from a low-resolution sequence. Eren et al. [EST97] extended the algorithm by using a validity map to disable projections based on observations with inaccurate motion information in the presence of motion estimation errors, as well as using the segmentation map to enable object-based processing where more accurate motion models can be utilized to improve the quality of the reconstructed image. Similarly, Shah and Zakhor [SZ99] also proposed an algorithm that specifically accounted for the possibility of inaccurate motion estimation, and compensated for the inaccuracies in multi-frame super-resolution.

Addressing face images, Baker and Kanade [BK99] presented super-resolution optical flow which took as input a conventional video stream, and simultaneously computed both optical flow and a super-resolution version of the entire video.

### 2.1.3 Learning-based super-resolution

Most learning-based super-resolution algorithms are based on statistical machine learning and work by analyzing large data sets of images of a particular class, for example faces, and learning the mapping from low-resolution to high-resolution images of that class. This enables them to infer, for example, the most likely high-resolution image depicting the same object as a low-resolution image given as input. When applied to face images, this process is known as face hallucination, first proposed by Baker and Kanade [BK00, BK02], and has been an active research area at the intersection of computer vision and computer graphics for the last decade [Ded07, FPC00, FJP02, LLX01+, LSZ01, LSF07, WT03, WT05]

Baker and Kanade [BK00, BK02] first developed a face hallucination method based on a prior on the spatial distribution of the image gradient for frontal face images. It infers the high-frequency component from a parent structure by recognizing the local features of the training set, and aims to generate extremely high-quality HR images of human faces from LR images. For example, given a LR image of 12 × 16 pixels only, which could barely be recognized as a face, face hallucination can synthesize an HR image of 96 × 128 pixels. Their algorithm can yield 4 to
8 fold improvements in resolution.

Liu et al. [LSZ01, LSF07, LLX+01] introduce a two-step statistical modeling approach that integrates both a global parametric model and a local non-parametric model. The first step is to derive a global linear model to learn the relationship between HR face images and the corresponding smoothed down-sampled LR ones. The second step is to model the residue between an original HR image and the reconstructed HR image by a non-parametric Markov network. Then by integrating both global and local models, they generate the photo-realistic face images.

Instead of using a probabilistic model as in Liu’s work, Wang et al. [WT05, WT03] propose a face hallucination model using PCA to represent the structural similarity of face images. They render the new hallucinated face image by mapping between the LR and HR training pairs of face images. In the PCA representation, different frequency components are independent. By selecting the number of eigenfaces, they extract the maximum amount of facial information from the low-resolution face image and remove the noise.

Inspired by the fact that belief propagation converges quickly to a solution of the Markov network, Freeman et al. [FJP02, FPC00] explore a simpler and faster one-pass algorithm, which uses the same local relationship information as the Markov networks but requires only a nearest-neighbor search in the training set for a vector derived from each patch of local image data. Their algorithms are an instance of a general-training-based approach that can be useful for image processing or graphics applications. It can be applied to enlarge images, remove noise, and estimate 3D surface shapes.

2.2 Super-resolution for medical imaging

Because of the benefits of image super-resolution, several methods have been proposed for the purpose of medical image super-resolution. Irani et al. [IP91] proposed iterative back-projection (IBP) to improve image resolution. This algorithm is feasible for monochrome and color image sequences when the relative displacements can be computed, and with approximate knowledge of the imaging process. Greenspan et al. [GOKP02] applied this iterative superresolu-
tion algorithm to construct high-resolution magnetic resonance images (MRI). Kennedy et al. [KIF\textsuperscript{+}06, KIF\textsuperscript{+}07] successfully applied this algorithm to construct positron emission tomography (PET). Hsu et al. [HYL\textsuperscript{+}04] proposed a study of applying wavelet-based projection-onto-convex-set super-resolution reconstruction algorithm to enhance spatial resolution of MRI heart images from a temporal sequence. Lettington and Hong [LH95] proposed an efficient ringing reduction algorithm which reduces the ringing artifacts that arise from the use of the unconstrained Poisson MAP method without much increased computing complexity. Kanemura et al. addressed the hyperparameter estimation problem in Bayesian image superresolution with a compound Gaussian Markov random field (MRF) prior [KMI07]. They estimated all the hyperparameters, the registration parameters, and the high-resolution image by means of minimizing variational free energy under the assumption of a factorized posterior. Nguyen et al. [NAL05] presented a new and efficient wavelet-based algorithm for image super-resolution, which is a combination of interpolation and restoration processes that exploits the interlaced sampling structure in the low resolution data.

### 2.3 Super-resolution for 3D face scanning

There has been a lot of work on high-resolution facial scanning. Existing methods are based on laser scanning, or structured light scan, or photometric stereo.

Laser-based scanning have been used for decades. Laser scanning projects a laser dot or line onto an object, a sensor measures the distance to the surface, then data is collected and recorded as point clouds which can be converted into a triangulated mesh. One advantage of laser scanning is that scanning device can be hand-held and portable. However, laser scanning is not suitable for translucent objects, because the projected laser line will be blurred due to subsurface scattering, and then the 3D models will also be blurred. As human skin is semi-translucent, using laser scanning for high-resolution facial scanning need to make a plaster copy of the human face first. However this process is time consuming, uncomfortable, and needs additional efforts to make the plaster copies.
Structured light scanning projects a pattern of light on the subject, and looks at the deformation of the pattern on the subject, and then uses a technique similar to triangulation to calculate the distance of every point on the line [RHHL02, ZSCS04, DRR05, ZH06]. These methods can achieve real-time processing, but they fail to provide high-resolution details for facial scanning.

Photometric stereo [Woo80, Woo89, WGT+05, MWGA06, HVB+07] estimates the surface normals of objects by observing that object under different lighting conditions. The technique was originally introduced by Woodham [Woo80]. It determines surface orientation of a Lambertian surface from its appearance under different lighting directions using a linear system. Nayar et al. [NIK90] use an extraction algorithm to compute orientation as well as relative strengths of the Lambertian and specular reflection components using a set of image intensity values measured at each surface point. Their method enables photometric stereo to adapt to variations in reflectance properties from one scene point to another. Schlüns et al. [SW93] extended photometric stereo to non-Lambertian surfaces using a two-stage process without introducing additional light sources or assuming a known micro facet distribution. Mallick et al. [MZKB05] present a photometric stereo method for non-diffuse materials that does not require an explicit reflectance model or reference object. Another line of work estimates surface orientation specifically from its specular reflection. Ikeuchi [Ike81] uses a distributed light source from uneven illumination of a fluorescent tube lights reflecting onto diffuse surfaces to estimate surface orientation. Halstead et al. [HBKM96] present an algorithm that reconstructs the shape of cornea from a single videokeratoscope image. They fit a surface to a set of normals rather than to a set of positions, and use specular reflection of a pattern of rings and spline-based surface fitting to interactively visualize the cornea. Recently, Wenger et al. [WGT+05] used time-multiplexed and high-speed cameras to capture time-varying reflectance properties, such as surface normals, albedo and ambient occlusions, of a live performance. Hernandez et al. [HVB+07] capture and track 3D moving data in real-time using multi-spectral photometric stereo. Malzbender et al. [MWGA06] use photometric stereo to recover per-pixel estimates of surface orientation, and then transform the reflectance based on recovered normal directions to enhance the surface detail. However they require high-speed camera and fast GPU to achieve real-time processing, and they don’t account
for specularities and shadows.

Ma [MHP+07] introduced a technique for high-resolution face scanning that combining structured light and photometric stereo scanning. They first obtain base geometry through structured light scan, and then capture photometric surface normal from either diffuse or specular reflectance using four spherical gradient illumination patterns. The surface normals are later used to optimize the base geometry to recover detailed facial features using an embossing process as in [NRDR05]. Golovinskiy et al. [GMP+06] introduced a new statistical technique for the analysis and synthesis of small detailed facial features such as wrinkles and pores, based on the analysis of high-resolution face scans.

In this thesis, I go further than previous work by Ma [MHP+07] for acquiring specular normal map by synthesizing the specular normals from diffuse normals instead of capturing both of them, and then generate detailed facial geometry.

### 2.4 Chapter summary

Reconstruction-based super-resolution is fast and in general improves the resolution of the input image, but the high-frequency components of the enhanced images are not reconstructed very well. In addition, reconstruction-based super-resolution provide less and less useful information as the magnification factor increases, and for large enough magnification factors any smoothness prior leads to overly smooth results with very little high-frequency content, no matter how many low-resolution input images are used.

Learning-based super-resolution breaks this limitation of reconstruction-based super-resolution using statistical machine learning. Aiming at analyzing large data sets of images of a particular class of objects and learning the mapping from low-quality to high-quality images of that class, learning-based super-resolution infers significantly better high-resolution images as a low-resolution image given as input. However learning-based super-resolution requires a very long time to perform the task. The time for training required by the algorithms is proportional to the size of the training dataset, so that the learning process is very time-consuming.
Additionally, some details in the super-resolution images are different from the original images. For some applications, for example medical imaging, it is necessary to further determine whether the differences would affect clinical diagnostic performance.

Furthermore, although learning-based super-resolution can take a degraded image or a sequence of images as input and produce an image or a sequence of images of higher quality, the definition of degraded images is subjective and mostly depends on the applications. The degraded images can be low-resolution surveillance videos, or mammograms obtained with very low dosage of radiation, or even compressed images with very high compression ratio. The concept of image quality can also be defined in various ways. For example, in photography, quality refers to the appreciation of a human observer, while in object detection tasks quality may refer to good localization of edges. Based on the definition of degraded image and image quality, the super-resolution algorithms as well as the results may greatly vary. For example, though the SR images have very high-resolution, some details in the super-resolution images are generally different from the original images. For medical imaging, it is necessary to further determine whether the differences would affect clinical diagnostic performance.

The question addressed in this dissertation is whether super-resolution algorithms can be tailored to the specific requirements of the application for which the images are used. In this thesis I develop super-resolution algorithms that can recover degraded images according to specific requirements of the application for which the images are used.
Chapter 3

A Stochastic Method for Face Image Super-Resolution

3.1 Introduction

In digital image analysis applications, high-resolution (HR) images are often required. However, in surveillance systems, the regions of interest are often impoverished or blurred due to the large distance between the camera and the objects, or the low spatial resolution of the sensing devices. Figure 3.1 illustrates an image collected from a surveillance video. In this image, people at a large distance appear very small and their faces cover a small number of pixels, without enough detail to enable analysis by humans or automated face recognition programs. In these type of applications, a way to enhance these low-resolution (LR) images is needed.

Figure 3.1: Faces in surveillance images

Image super-resolution should provide an improvement in the perceived detail content compared to that of the original images. This typically involves restoration of the high-frequency
content, which in turn requires an increase in pixel density. Furthermore, in the process of capturing digital images, several problems can affect the quality of the sensed images, such as CCD variations due to different responses of different cells to identical light intensities, scattering due to the medium through which the light beams pass, motion blur due to limited shutter speed, and quantization effects. Hence, the images obtained from digital cameras are distorted. Thus, image super-resolution is closely related to image restoration, which aims to enhance a degraded image without changing its size.

I introduce a fast object-specific super-resolution approach. The approach combines separate global and local modeling stages. Global modeling, which provides the general structure of the face image, is done by eigentransformation [WT03, WT05], while local modeling, which provides high frequency details, is performed using a novel stochastic search algorithm that efficiently finds near-optimal local patches in a training set of face images. The experimental results show that this approach provides high-quality results while significantly reducing running times relative to other state-of-the-art methods.

3.2 Framework

As shown in Figure 3.2, the proposed method consists of two steps. The first step uses eigentransformation to infer global faces $I^g_H$, which is named global modeling. Principal Component Analysis (PCA) is used to fit the input face images as a linear combination of the LR face images in the training set. The HR images are then inferred by replacing the LR training images with HR ones, while retaining the same combination coefficients [WT05]. At the second step, high-frequency contents of the HR images, $I^l_H$, are captured by a patch-based one-pass algorithm, [FJP02], which is named local modeling. To improve the efficiency of searching the most compatible patch, I introduce stochastic search into the one-pass algorithm, which is a probabilistic method that iteratively propagates the targets’ position using Bayes’ rule. Finally, the super-resolution image, $I_H$, is the sum of global face and local face, $I_H = I^g_H + I^l_H$. The details of these algorithms are given in the next three sections.
3.3 Global modeling

In global modeling I use an algorithm originally introduced by Wang [WT05], which is called eigentransformation. The eigentransformation is a simple and powerful technique for image enhancement based on principal component analysis (PCA). It assumes that there is a training set of pairs of images \( \langle (L_1, H_1), \ldots, (L_n, H_n) \rangle \), where each pair \((L_i, H_i)\) contains a low resolution face image \(L_i\) and its corresponding high-resolution counterpart \(H_i\). The eigentransformation allows to represent any image as a linear combination of images in the training set. When given a low resolution image \(L\), it finds the vector of coefficients \([c_1, \ldots, c_n]\) so that

\[
L = \sum_{i=1}^{n} c_i L_i + \mu_L
\]

where \(\mu_L\) is the mean low-resolution face.

Given the vector \([c_1, \ldots, c_n]\), the approximate high resolution image \(H\) can be computed by

\[
H = \sum_{i=1}^{n} c_i H_i + \mu_H
\]

where \(\mu_H\) is the mean high-resolution face image.

Because the coefficients are not computed from the HR training data, some non-face-like distortion may be introduced. To reduce the distortion, I apply constraints by bounding the projection onto each eigenvector by its corresponding eigenvalue, then the synthesized face image is reconstructed from these constrained coefficients.
3.4 Local modeling

Given a global face, to construct the corresponding local face, I first filter the global face with a Gaussian high-pass filter, and then subdivide the filtered global face into patches, which is called the low-frequency patches of the HR faces, by scanning a window across the image in raster-scan order. Similarly, I also filter and subdivide the HR faces in the training set into patches which is called high-frequency patches of the training HR faces.

To construct a local face, for each low-frequency patch, a high-frequency patch of the training HR face is selected by a nearest neighbor search from the training set based on local low-frequency details and adjacent, previously determined HR patches. The selected high-frequency patch should not only come from a location in the training images that has a similar corresponding low-frequency appearance, but it should also match at the edges of the patch with the overlapping pixels, which is called high-frequency overlap, of its previously determined high-frequency neighbors to ensure that the high-frequency patches are compatible with those of the neighboring high-frequency patches.

In this thesis I compute the local faces with an algorithm that is an extension of the one-pass algorithm, proposed by Freeman et al. [FJP02, FPC00]. In the one pass algorithm, I first concatenate the pixels in the low-frequency patch and the high-frequency overlap to form a search vector. The training set also contains a set of such vectors. Then I search for a match by finding the nearest neighbor in the training set. When a match is found, I extract the corresponding high-frequency patch from training data set and add it to the initial global face to obtain the output image.

Mathematically, this process can be described as follows. Suppose there is a training data set

\[ \{(x^{(i,j,k)}, y^{(i,j,k)}, z^{(i,j,k)})\}, \]

\[ i = 1, 2, \ldots, l; j = 1, 2, \ldots, m; k = 1, 2, \ldots, n \}

where \( x^{(i,j,k)} \) is the low-frequency patch at the \( i^{th} \) row and \( j^{th} \) column of the \( k^{th} \) training HR face image, \( y^{(i,j,k)} \) is the corresponding high-frequency overlap and \( z^{(i,j,k)} \) is the corresponding
high-frequency patch of the training HR face image, \( l \) is the number of rows of patches in a training image, \( m \) is the number of columns of patches in a training image and \( n \) is the number of training images.

Given an input LR patch \( \overline{x} \), I need to find an HR patch \( z(i', j', k') \) such that

\[
(i', j', k') = \arg \min_{i,j,k} (d(\overline{x}, x^{(i,j,k)}) + \alpha \times (d(y^{(i,j,k)}, y^{(i,j,k)}_N)))
\]

where \( d(x, y) \) is the Euclidean distance between \( x \) and \( y \), \( y^{(i,j,k)}_N \) is the overlap of \( z^{(i,j,k)} \) with the adjacent, previously determined high-frequency patches \( z^{(i-1,j,k)} \) and \( z^{(i,j-1,k)} \), \( \alpha \) is a user-controlled weighting factor, and \( z(i', j', k') \) is the selected high-frequency patch.

### 3.5 Stochastic search in local modeling

![Figure 3.3: Local modeling with stochastic search](image)

This thesis improves the efficiency of face image super-resolution using stochastic search for local modeling by exploiting the fact that face patches maintain relatively tight distributions for shape at successive iterations. For example, suppose I found the best match at iteration \( t \), which is a patch from a left eye taken from training image \( T \); intuitively, there is a high probability that the best match at iteration \( t + 1 \) should also belong to a left eye and come also from \( T \), or from a training image that is similar to \( T \). Therefore the position of the most compatible high-frequency
patch at iteration \( t \), \( z_t \), and its history \( Z_t = z_1, z_2, \ldots, z_{t-1} \) form a temporal Markov chain, so that the new position is conditioned directly only on the immediately preceding state, independent of its earlier history.

\[
P(z_t|z_t) = P(z_t|z_{t-1})
\]

Therefore, based on \( z_{t-1} \), I could estimate the most likely positions of \( z_t \) and I just need to search these positions instead of the exhaustive search performed by the conventional one-pass algorithm.

The key idea of the proposed algorithm is shown graphically in Figure 3.3, where I show how a stochastic search is performed using candidate patches taken from locations and images that are similar to the most recently found patch.

Let \( z^{(i,j,k)}_t \) be the most compatible patch found at iteration \( t \) (for \( t = 0 \), an exhaustive search needs to be performed). To find the best patch at iteration \( t + 1 \), we generate a set of \( N \) candidate patch locations \( \{(i', j', k')^1, \ldots, (i', j', k')^N\} \) around location \((i, j)\) in images that are similar to image \( k \).

Each location \((i', j', k')^q\) will be randomly generated according to the following distributions:

\[
\begin{align*}
    i' &= i + \alpha \\
    j' &= j + \beta \\
    k' &= \gamma(A(k))
\end{align*}
\]

where \( \alpha \) and \( \beta \) are normally distributed random variables, \( A(k) \) is a list of face images that are similar to image \( k \) (which includes \( k \) itself) and \( \gamma(.) \) is a sampling function that randomly selects face images from \( A(k) \) with a probability that is directly proportional to their similarity with face image \( k \). In a preprocessing stage, I built a directed graph where every vertex contains a face image and there is an edge \( e(k, v) \) if vertices \( k \) and \( v \) contain similar face images, according to a mean-squared distance metric, as illustrated in Figure 3.4. Thus \( A(k) \) is the set of vertices that are adjacent to \( k \) in this graph.
After generating the \( N \) candidate patches, I select as the next patch the best match within the generated point set and repeat the process again.

The proposed algorithm generates a set of \( N \) candidates around the point \((i,j,k)\). Suppose I have \( n \) images in the training data set and each image has \( k \) patches, for exhaustive search, the running time for constructing one local face is \( O(nk^2) \); for stochastic search, the running time is \( O(kn + kN) \).

![Figure 3.4: Similar faces are connected in a graph structure](image)

### 3.6 Experimental results

For experiments I use the BioID data set which consists of 1521 gray-level face images of \( 384 \times 288 \) pixels with a frontal view of 23 different people under a high variety of lighting conditions, backgrounds and face sizes. As the variations of human faces, such as glasses and beard, may greatly affect the performance of face hallucination in global modeling, I construct class-based subsets that contain images of small variations as training dataset. In preprocessing I crop and normalize, and then register the face images by 23 manually selected facial feature points (Figure 3.5), so that I can assume that the same parts of faces appear in roughly the same parts of the images. The face image size is fixed to \( 128 \times 128 \) pixels and I use them as the training HR face images. The training LR face images are down-sampled from the training HR face images by averaging the neighborhood pixels. In our experiments, the down-sample factor is 8.
To construct the training dataset, I generate global faces from the training LR face images and filter them with a Gaussian high-pass filter. Then I subdivide the filtered global face images into low-frequency patches by scanning a $4 \times 4$ pixel window across the image in raster-scan order. Then I again filter and subdivide the training HR face images into $4 \times 4$ pixel high-frequency patches. At each step I also get a 9-bit overlap of each high-frequency patch with the high-frequency patches above and to the left. Then I create the training vectors by concatenating the low-frequency patches and corresponding high-frequency overlaps. In practice, the size of low-frequency patches and high-frequency patches is not necessarily the same. The parameter $\alpha$, which controls the trade-off between matching the low-frequency patches and finding the most compatible high-frequency patches, is set to 0.2.

Figure 3.6 shows the face image super-resolution results of exhaustive search. The LR inputs are $16 \times 16$ pixel face images. The super-resolution face images are $128 \times 128$ pixel face images. Compared with the input images and the bicubic interpolation results, the super-resolution face images have much clearer features.

Figure 3.7 to 3.10 show the examples of face image super-resolution results of stochastic search. In Figure 3.7, the size of the randomly generated search set is 300, which is too small, so that the quality of local face is low. In Figure 3.8, the size of the randomly generated search set is 1200. In Figure 3.9, the size of the randomly generated search set is 500. In Figure 3.10, the size
of the randomly generated search set is 2000. The results show that a larger generated search set provides a better-quality local face.

In Table 3.1, I report the Mean-Squared-Error (MSE) between the super-resolution face images and the ground truth. Although the MSE is a physically meaningful metric for signal reconstruction, it does not necessarily reflect perceived visual quality by humans [Gir93]. Thus I also use the mean Structural Similarity (SSIM) Index, which aims to primarily measure the structural changes between a reference image and its distorted version [WBSS04]. These results confirmed the results observed from Figure 3.7 through 3.10 that generally a better-quality local face can be generated with a larger search set.

In Table 3.2 I report the computation time testing results. The time to synthesize a local face of $128 \times 128$ pixels using the stochastic search, with a randomly generated set of size 2000, is about 16s, while using the exhaustive search takes about 99s, on a 2.4 GHz PC in Matlab.

From Table 3.1 and 3.2, and Figure 3.7 to 3.10, I can see that stochastic local modeling with a larger generated search set generates better-quality local faces but takes a longer time. Therefore, a tradeoff between quality and efficiency has to be considered.

\[
\begin{array}{|c|c|c|}
\hline
N & MSE & SSIM \\
\hline
300 & 37.2095 & 0.9319 \\
500 & 30.1520 & 0.9387 \\
1200 & 28.3502 & 0.9386 \\
2000 & 29.6137 & 0.9396 \\
\text{exhaustive} & 31.2918 & 0.9403 \\
\hline
\end{array}
\]

Table 3.1: Quantitative evaluation
3.7 Chapter summary

In this chapter, I presented a framework for face image super-resolution integrating global modeling and local modeling. In global modeling, I infer the global face of the input LR face image with a linear combination of PCAs from the HR faces in the training set. In local modeling, I use a stochastic patch-based one-pass algorithm to infer the local face. The final super-resolution image is the sum of the global face and the local face. The most computational intensive part of this approach is local modeling, the computation time of which is proportional to the size of the training dataset. I introduce a stochastic local search into the one-pass algorithm that constraints the search space to a fixed size and makes real-time super-resolution possible. If there are $n$ images in the training data set, each image has $k$ patches, and the generated set contains $N$ patches, this method reduces the running time for constructing one local face from $O(nk^2)$ in exhaustive search to $O(kn + kN)$ in stochastic search. Experimental results show that the difference in quality relative to exhaustive search is negligible.
Figure 3.6: (a) Original high-resolution face images. (b) Input low-resolution face images. (c) Generated super-resolution face images. (d) Results of bicubic interpolation. (e) Global faces. (f) Local faces. This figure shows the super-resolution results of exhaustive search. Compared with the input low-resolution images and the bicubic interpolation results, the super-resolution face images have much clearer detailed features.
Figure 3.7: (a) Original high-resolution face images. (b) Input low-resolution face images. (c) Generated super-resolution face images. (d) Results of bicubic interpolation. (e) Global faces. (f) Local faces. This figure shows the results of super-resolution using stochastic search. The size of randomly generated search set is set to 300. However, this is too small, so that the quality of local face decreases.

Figure 3.8: (a) Original high-resolution face images. (b) Input low-resolution face images. (c) Generated super-resolution face images. (d) Results of bicubic interpolation. (e) Global faces. (f) Local faces. In this figure, the size of randomly generated search set is set to 1200.
Figure 3.9: (a) Original high-resolution face images. (b) Input low-resolution face images. (c) Generated super-resolution face images. (d) Results of bicubic interpolation. (e) Global faces. (f) Local faces. In this figure the size of randomly generated search set is set to 500.
Figure 3.10: (a) Original high-resolution face images. (b) Input low-resolution face images. (c) Generated super-resolution face images. (d) Results of bicubic interpolation. (e) Global faces. (f) Local faces. In this figure, the size of randomly generated search set is set to 2000. From Figure 3.7 to 3.10, we can see that stochastic local modeling with a larger generated search set would have better quality of local faces but take a longer time. Therefore, a tradeoff between quality and efficiency need to be found.
Chapter 4

Patch-based Simultaneous Face Super-resolution and Recognition

4.1 Introduction

Face recognition from surveillance images is a difficult and important problem. Since surveillance cameras are usually set with large fields of view in order to capture as much of the scene as possible, objects of interest, such as faces, are commonly a few pixels in size. State-of-the-art face identification systems work accurately with face images of at least $20 \times 20$ pixels, thus their application in surveillance is severely limited. Some methods have been proposed to address the problem of recognition from low-resolution (LR) images, but recognition rates are still low [GBA+03].

Recently, several methods for increasing the resolution of face images, collectively as “face hallucination”, have been proposed [BK00, BK02, WT05, LSZ01, LSF07, LLX+01, FJP02, FPC00]. These algorithms employ machine learning techniques to approximate the mapping from low resolution to high resolution images and produce highly realistic images. However, though face super-resolution (SR) can improve the appearance of face images dramatically, the detailed facial features such as eyes, eyebrows, nose, mouth and teeth of the hallucinated face are different from the ground truth. Further study of whether they help face recognition, by either human beings or computers, is still needed.

In psychology, pioneering work on face identification by human beings with low-resolution imagery was done by Harmon and Julesz [HJ73] and Morrone et al. [MBR83]. Working with low-resolution images of familiar faces as shown in figure 4.1, they found that human beings
can achieve high recognition accuracies, approximately 95%, even with images containing just $16 \times 16$ pixels. Furthermore, Costen et al. [CPC93] presented data that shows the dependence of face identification performance by human beings on face images with systematically varied resolutions.

![Figure 4.1: Low-resolution face images](image)

Given the significant improvement of the face appearance by the face super-resolution process, it is interesting to investigate whether the super-resolution helps automatic face identification. In this thesis, I study whether face SR can help automatic face identification. I applied a standard face hallucination algorithm as a preprocessing stage prior to recognition and also propose a method that instead of performing SR and identification as two separate sequential processes, integrates the two tasks together by directly computing a maximum probability identity parameter from SR for identification based on local patterns of human faces. The experiments show that face SR did not improve the performance of conventional face identification approaches, however, the method that combines face SR and face identification significantly improves recognition accuracy.

### 4.2 Simultaneous face super-resolution and identification

For patch-based simultaneous face identification, I first construct a dataset of multi-person face images. The dataset can be expressed as
\[ D = f_1 \cup f_2 \cup f_3 \cup \ldots \cup f_T \]

where \( D \) is the dataset of multi-person face images, \( f_1 \) is the face images of one specific person, and so on. As I did in face super-resolution in chapter 3, I then decompose the training set into low-frequency patches and the high-frequency overlaps as follows

\[
\{(x^{(i,j,k,t)}, y^{(i,j,k,t)}, z^{(i,j,k,t)}),
\]

\[
i = 1, 2, \ldots, l; j = 1, 2, \ldots, m; k = 1, 2, \ldots, n; t = 1, 2, \ldots, T\}
\]

where \( x^{(i,j,k,t)} \) is the low-frequency patch at the \( i^{th} \) row and \( j^{th} \) column of the \( k^{th} \) training HR face image of the \( t^{th} \) person, \( y^{(i,j,k,t)} \) is the corresponding high-frequency overlap. \( z^{(i,j,k,t)} \) is the corresponding high-frequency patch of the training HR face image, \( l \) is the number of rows of patches in a training image, \( m \) is the number of columns of patches in a training image and \( n \) is the number of training images, \( T \) is the total number of people.

When searching the high-frequency best matching patch in local modeling, I record the probability identity parameter of the patch in table \( w \). The probability identity parameter can be expressed as

\[
w(i, j, k, t) = \max_{i,j,k,t} (1/(d(x, x^{(i,j,k,t)})) + \alpha \ast (d(y^{(i,j,k,t)}, y_N^{(i,j,k,t)})))
\]

where \( d(x, y) \) is the Euclidean distance between \( x \) and \( y \), \( y_N^{(i,j,k,t)} \) is the overlap of \( z^{(i,j,k,t)} \) with the adjacent, previously determined high-frequency patches \( z^{(i-1,j,k,t)} \) and \( z^{(i-1,j,k,t)} \), \( \alpha \) is a user-controlled weighting factor, and then \( H \) is the identification result.

\[
H(\gamma) = \arg \max_{i,j,k,t} \sum_{i,j,k,t} w_{i,j,k,t} I(t = \gamma)
\]

where \( I(.) \) is the indicator function that takes 1 if the argument is true and 0 otherwise.
4.3 Experiments

I compare face identification results in different resolutions (128 × 128, 64 × 64, 32 × 32, 16 × 16, 8 × 8, 4 × 4). Two algorithms are used for face identification: using neural networks with Principal Components (PCs) and patch-based face identification. For testing, 10-fold cross validation is used.

AT&T dataset is used to test the algorithms. The AT&T dataset consists of 360 face images from 40 people for training, and 40 face images from 40 people for testing.

The neural network approach uses principal components (PCs) as the features. Because in images of different resolutions the same number of principal components will contain different amount of information, I also test neural networks with different number of principal components.
in different resolutions that contain about 90% of facial information. The neural network consists of two layers, using as transfer functions. The maximum number of training iteration is 1000. The minimum error is set 0.0005.

The patch-based approach consists of two steps. The first step uses eigentransformation to infer global faces. Principal Component Analysis (PCA) is used to fit the input face images as a linear combination of the low resolution face images in the training set. The HR images are then inferred by replacing the LR training images with HR ones, while retaining the same combination coefficients. At the second step, a patch-based one-pass algorithm captures high-frequency contents of the HR images. Meanwhile, HR patch compatibilities for neighboring HR patches that are already selected, and LR similarities between the selected patch in training dataset and the corresponding input patch, are computed and stored for the task of face identification.

To construct the training dataset, I generate global faces from the training LR face images and filter them with a Gaussian high-pass filter. Then I subdivide the filtered global face images into low-frequency patches by scanning a $4 \times 4$ pixel window across the image in raster-scan order. Then I again filter and subdivide the training HR face images into $4 \times 4$ pixel high-frequency patches. At each step I also get a 9-bit overlap of each high-frequency patch with the high-frequency patches above and to the left. Then I create the training vectors by concatenating the low-frequency patches and corresponding high-frequency overlaps. In practice, the size of low-frequency patches and high-frequency patches is not necessarily the same. The parameter $\alpha$, which controls the trade-off between matching the low-frequency patches and finding the most compatible high-frequency patches, is set to 0.2.

Figure 4.6 gives the face identification results on AT&T dataset using patch-based, and for comparison it also reports identification results adopting neural network approach, on different resolutions. The results show that patch-based face identification can help automatic face identification, specially in low-resolution images.

Figure 4.3 through Figure 4.5 reports the face image super-resolution results. The LR inputs are $4 \times 4$, $16 \times 16$, and $32 \times 32$ pixel face images. The super-resolution face images are $128 \times 128$ pixel face images. Compared with the input images, the super-resolution face images have much
clearer facial features.

Figure 4.7 reports the experimental results of face identification on super-resolution face images using neural network approach, on different resolutions. When performing face identification, I first enhance the resolution of low-resolution images using super-resolution algorithms to 128x128 pixels, then use neural networks for identification. For comparison, I also report the neural network approach with bilinearly interpolated images. The results show that interpolation has little to do with the identification tasks, and face super-resolution doesn’t really help face identification using neural network.

4.4 Chapter summary

In this chapter, I applied a standard face hallucination algorithm as a preprocessing stage prior to recognition and also propose a method that instead of performing super-resolution and identification as two separate sequential processes, integrates the two tasks together by directly computing a maximum probability identity parameter from super-resolution for identification based on local patterns of human faces. The experiments show that face super-resolution used as a preprocessing stage did not improve the performance of conventional face identification approaches, however, the method that combines face super-resolution enhancement and face identification significantly improves recognition accuracy.

Compared with conventional face recognition algorithms, the drawback of our approach is that it requires very long time to perform the recognition task, as the time required by the algorithm is proportional to the size of the training dataset. As future work, to reduce the running time of the super-resolution algorithm, I will use stream processing to parallelize the execution. The proposed method only uses information derived from the local model; I am currently working on ways of taking advantage of the information that can be obtained from the eigenface-domain in the global model.
Figure 4.3: Super-resolution results (a) Super-resolution face. (b) Global face. (c) Input low-resolution face image. (d) Original high-resolution face image.
Figure 4.4: Super-resolution results (a) Super-resolution face. (b) Global face. (c) Input low-resolution face image. (d) Original high-resolution face image.
Figure 4.5: Super-resolution results (a) Super-resolution face. (b) Global face. (c) Input low-resolution face image. (d) Original high-resolution face image.
Figure 4.6: Face identification results of 10-fold cross validation on AT&T dataset

Figure 4.7: Face identification results of low-resolution images, interpolated images and super-resolution images on AT&T dataset using neural networks
Chapter 5

Super-resolution of Mammograms

5.1 Introduction

Breast cancer causes more deaths than any other type of cancer for females worldwide. However, if breast cancer can be detected early, the five-year survival rate increases considerably [JSW+09]. High-quality mammography, which consists of using X-rays to examine the human breast, is the most effective technology presently available for breast cancer screening.

The quality of diagnoses depends on the resolution of the mammograms, among other factors, with higher resolutions providing a higher level of detail that normally leads to improved accuracy. However, to obtain higher resolutions, larger doses of radiation are necessary, which may have harmful effects for patients. To alleviate this problem, some work has been done attempting to increase the resolution of mammograms without a corresponding increase in radiation. Robinson et al. [RFL+07] used multi-frame image reconstruction to produce high-resolution mammograms beyond the native resolution of a digital image sensor by way of accurate sub-pixel registration of aliased images. Several other image enhancement techniques have been proposed for this problem [IP91, GOKP02, KIF+06, KIF+07, KMI07, WS09]; however, improvements have been modest.

Recent research at the intersection of computer vision and computer graphics has produced methods for automatically increasing the resolution of images of specific classes, commonly faces [BK02, WT03]. These methods use statistical machine learning techniques to learn the function that maps low resolution images to their high resolution counterparts, for a particular class of images. Efforts to date in this line of work have focused on face images; one of the goals of this thesis is to investigate whether these methods can be extended to other domains, particularly
medical image processing.

In this chapter, I present an end-to-end method to synthesize a high-resolution (HR) mammogram given a low-resolution (LR) one as input. This method receives as input low-resolution mammograms, which can be generated with low doses of X-ray radiation. Then these images are automatically registered and aligned using a mesh warping algorithm. Finally, a two-step super-resolution algorithm is applied, which integrates a PCA-based global model and a patch-based local model, to generate the high-resolution images. The experimental results show that the algorithm can generate high-quality high-resolution mammograms from low-resolution input with no manual registration.

5.2 Related work

Because of the benefits of image super-resolution, several methods have also been proposed for the purpose of medical image super-resolution. Irani et al. [IP91] proposed an iterative algorithm together with a method for image registration with subpixel accuracy to increase image resolution. The approach is similar to reconstruction of a 2-D object from its 1-D projections in computer aided tomography (CAT). Images are reconstructed from their projections in many directions in tomography, while in super-resolution, each low-resolution pixel is a projection of a region in the scene whose size is determined by the imaging blur. The high-resolution image is then constructed using an approach similar to the back-projection method used in CAT.

Based on Irani’s work, Greenspan et al. [GOKP02] applied the iterative super-resolution algorithm to construct high-resolution magnetic resonance images (MRI). MRI slice thickness is determined by hardware limitations and pulse sequence timing considerations. Thus the in-plane resolution is higher than the resolution of the slice-select direction. They addressed the challenge of achieving HR isotropic 3-D MRI images by merging several sets of 2-D slices in slice-select direction.

Kennedy et al. [KIF+06, KIF+07] successfully applied Irani’s iterative super-resolution algorithm to construct positron emission tomographies (PET). PET resolution is limited by physi-
cal parameters such as scatter, counting statistics, positron range, patient motion, detector array geometry, and the implemented acquisition protocol. They demonstrated how an iterative super-resolution algorithm can be implemented to improve PET resolution using shifts and rotations in the trans-axial plane as well as along the axial direction.

Hsu et al. [HYL+04] proposed a wavelet-based projection-onto-convex-set super-resolution reconstruction algorithm to enhance spatial resolution of MRI heart images from a temporal sequence. This approach makes use of the non-stationary effect in the successive images in the sequence to extract information for image reconstruction at a higher spatial resolution.

Lettington and Hong [LH95] proposed an efficient algorithm to reduce the ringing artifacts that arise from the use of the unconstrained Poisson maximum a posteriori (MAP). They use a Lorentzian probability function to model the image by studying the distribution of its edge values, then introduced a correction term to increase the image likelihood using the mean square error (MSE) criterion. The correction term is effective in reducing the ringing artifacts while maintaining the sharpness of the image.

Kanemura et al. addressed the hyper-parameter estimation problem in Bayesian image super-resolution with a compound Gaussian Markov random field (MRF) prior [KMI07]. They estimated all the hyper-parameters, the registration parameters, and the HR image by means of minimizing variational free energy under the assumption of a factorized posterior.

Nguyen et al. [NAL05] presented an efficient wavelet-based algorithm for image super-resolution that is a combination of interpolation and restoration processes and exploits the interlaced sampling structure in the low resolution data.

5.3 Super-resolution algorithm

In this chapter I propose a mammogram super-resolution algorithm that can generate high-quality high-resolution images from low-resolution input with no manual registration. The proposed algorithm consists of four main steps. The first step automatically aligns the parts of the images containing the breasts to a standardized position. The second step uses eigentransformation to
infer global models, that is, the low-frequency components of the target image. Principal Component Analysis (PCA) is used to fit the input images as a linear combination of the low resolution images in the training set. The HR images are then inferred by replacing the LR training images with HR ones, while retaining the same combination coefficients. In the third step, a patch-based one-pass algorithm captures high-frequency contents of the HR images. The fourth step re-maps the breasts back to their original position. The details of eigentransformation and one-pass algorithm are described in chapter 3.

Image alignment is a key step for the success of the algorithm. In practice, we cannot assume that any low-resolution mammogram has been accurately aligned, although the approximate position of the breasts is given by mammography sensors. Therefore, in preprocessing, I automatically align the mammograms to make sure all breasts are in exactly the same position, and then do the super-resolution. The automatic alignment process consists of two parts, 2-pass mesh warping [Wol90] and segmentation-based initialization.

The 2-pass mesh warping algorithm accepts a source image and two 2-D arrays of coordinates. The first array, $S$, specifies the coordinates of control points in the source image, and the second array, $D$, specifies their corresponding positions in the destination image. The first pass is responsible for resampling each row independently. It maps all $(u,v)$ points to their $(x,v)$ coordinates in the intermediate image $I$, in which the x-coordinates are the same as those in $D$, and y-coordinates are same as those in $S$.

For each pixel $P$ in intermediate image $I$, the value of $P$ is evaluated as a weighted sum from $x_0$ to $x_1$, the left most and rightmost positions in $S$ that are the projections of the left and right integer-valued boundaries of $P$

$$P = \sum_{x=x_0}^{x_1} \frac{k_x S_x}{x_1 - x_0}$$

where $k_x$ is the scale factor of source pixel $S_x$, and the subscript $x$ denotes the index that lies between $\text{floor}(x_0)$ and $\text{ceil}(x_1)$. The scale factor $k_x$ is defined as
$$k_x = \begin{cases} 
[x] - x_0 & \text{if } \text{floor}(x) < x_0 \\
1 & \text{if } x_0 \leq x < x_1 \\
x_1 - [x] & \text{if } \text{ceil}(x) > x_1 
\end{cases}$$

The second pass then resamples each column in $I$, mapping every $(x, v)$ point to its final $(x, y)$ position. This is virtually identical to the first pass, we just need to substitute $(x, v)$ for $(u, v)$, and substitute $(x, y)$ for $(x, v)$ \[Wol90\].

The key to apply 2-pass mesh warping is to build the 2-D arrays of coordinates. We use segmentation-based initialization to build the 2-D arrays of coordinates automatically. The initialization for 2-pass mesh warping consists of the following steps (Figure 5.1):

1. Convert the input image to binary image.
2. Apply erosion to remove the labels in the binary image.
3. Convert the binary image to gradient image to find the edge.
4. Use skeletonization to reduce the edge to a single line.
5. Distribute points uniformly on each side of the line to generate the mesh for 2-pass mesh warping.

5.4 Experimental results

I use DDSM (Digital Database for Screening Mammography) for the experiments. DDSM is a standard dataset used by the mammography image analysis research community. The database has about 2,500 cases. Each case includes two images of each breast, along with some associated patient information and image information.

In this thesis, 400 normal left mediolateral oblique (MLO) images from the DDSM dataset are used for training and 10 normal left MLO images for testing. I use an automatically built $32 \times 8$ mesh to register the mammograms.
Figure 5.1: Automatic registration: (a) Original image. (b) Converting the original image to binary image. (c) Using erosion to remove the labels in the binary image. (d) Converting the binary image to gradient image to find the edge. (e) Using skeletonization to reduce the edge to a single line. (f) Sampling points on the line and making the mesh.

To construct the training dataset, I generate global models from the training LR mammograms and filter them with a Gaussian high-pass filter. Then I divide the filtered global model into low-frequency patches by scanning a 4 × 4 pixel window across the images in raster-scan order. Then I again filter and subdivide the training HR mammograms into 4 × 4 pixel high-frequency patches. At each step I also get a 9-pixel overlap of each high-frequency patch with the high-frequency patches above and to the left. Then I create the training vectors by concatenating the low-frequency patches and corresponding high-frequency overlaps. In practice, the size of low-frequency patches and high-frequency patches is not necessarily the same. The parameter $\alpha$, which controls the trade-off between matching the low-frequency patches and finding the most compatible high-frequency patches, is set to 0.2, which gives good HR results in the experiments.

To reduce the effect of background to the quantitative evaluation of the super-resolution results, I applied a mask to each image (Figure 5.3) to get the region of interests (ROI). The mask is computed automatically using an image segmentation algorithm. At the same time I removed the labels from the background.

For each high-resolution image, we downsample it by a factor of 2, 4, 8 and 16, and then enhance the low-resolution image to its original resolution. The experimental results of different resolutions are reported in Figures 5.2 to 5.5. I also measure the quality of super-resolution
results using PSNR and MSSIM (Table 5.1), which are commonly used image quality measures in most super-resolution studies [BF06]. The results indicate that the super-resolution images have much more high-frequency information than images obtained by nearest-neighbor and bilinear interpolation.

Comparing the different image quality measures, we can see that although the PSNR is a meaningful standard image quality measure [BF06], it does not necessarily reflect perceived visual quality by humans [Gir93][WBSS04]. In some of the experiments shown in Table 5.1, the PSNR value of SR by a downsample factor of 4 is 34.02, which is lower than the corresponding PSNR value of bilinear interpolation, 35.2363. But from Figure 5.3, we can see that the SR results have much more detail information than the results of bilinear interpolation. The MSSIM, which accounts well for the texture changes introduced by the super-resolution process, has its values increased by the super-resolution algorithm. Since this measure was created to better reflect perceived visual quality by humans [Gir93][WBSS04], this implies that the super-resolution algorithm increases perceptual quality when the downsample factor is considerable.

For comparison purposes, I also report in Table 5.2 the results using iterative back-projection (IBP) used by Greenspan et al. [GOKP02] and Kennedy et al. [KIF+06, KIF+07], and the phase-adaptive method using complex wavelets (PACW) tested by Wong et al. [WS09]. PSNR is used as image quality measure, and the downsample factor is 4 with respect to the HR mammograms. From Table 5.2 we can see that this method has noticeably higher values than the other methods.

### 5.5 Chapter summary

In this chapter I presented a new way to improve the quality of X-ray images without increased doses of radiation. The method takes low-resolution images obtained with low doses of X-ray radiation, and automatically registers and aligns them using mesh warping, and then uses super-resolution algorithms to create high resolution images from the LR input. My results show that the super-resolution algorithm can generate accurate high-resolution mammograms from low-resolution input with no manual registration.
<table>
<thead>
<tr>
<th>Method</th>
<th>Downsampling factor</th>
<th>PSNR</th>
<th>MSSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Super-resolution</td>
<td>16</td>
<td>35.7403</td>
<td>0.8095</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>34.7535</td>
<td>0.8238</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>34.0200</td>
<td>0.8242</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>33.4649</td>
<td>0.8223</td>
</tr>
<tr>
<td>Nearest-neighbor interpolation</td>
<td>16</td>
<td>31.5602</td>
<td>0.6164</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>32.9498</td>
<td>0.6690</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>34.4359</td>
<td>0.7600</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>36.6833</td>
<td>0.8558</td>
</tr>
<tr>
<td>Bilinear interpolation</td>
<td>16</td>
<td>32.3788</td>
<td>0.7282</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>33.9344</td>
<td>0.7577</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>35.2363</td>
<td>0.8072</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>37.1938</td>
<td>0.8807</td>
</tr>
</tbody>
</table>

Table 5.1: Image quality measures

As medical imaging moves towards complete digital imaging and produces prohibitively large amounts of data, compression is necessary for storage and communication purposes. This method has the potential to eliminate the need to store images at full resolution, since they could be re-generated from low-resolution ones. In next chapter, I will explore the trade-offs between compression ratios and accuracy, as well as analyze the interaction between this method and conventional compression schemes.

<table>
<thead>
<tr>
<th>Method</th>
<th>IBP</th>
<th>PACW</th>
<th>our method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downsampling factor</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>PSNR</td>
<td>29.91</td>
<td>32.51</td>
<td>34.02</td>
</tr>
</tbody>
</table>

Table 5.2: Comparison of PSNR of our methods with other methods
Figure 5.2: Sample super-resolution results with a downsample factor of 2: (a) Original high-resolution images. (b) Super-resolution images. (c) Interpolated images. (d) Low-resolution images.
Figure 5.3: Sample super-resolution results with a downsample factor of 4: (a) Original high-resolution images. (b) Super-resolution images. (c) Interpolated images. (d) Low-resolution images.
Figure 5.4: Sample super-resolution results with a downsample factor of 8: (a) Original high-resolution images. (b) Super-resolution images. (c) Interpolated images. (d) low-resolution images.
Figure 5.5: Sample super-resolution results with a downsample factor of 16: (a) Original high-resolution images. (b) Super-resolution images. (c) Interpolated images. (d) low-resolution images.
Chapter 6

Hybrid Compression of Mammograms
Using Super-resolution

6.1 Introduction

As mammography moves towards completely digital, technological advances in data storage and transmission have not kept up with the tremendous growth of digital data. This creates serious challenges for long-term storage and efficient transmission of mammograms. For example, a typical mammogram can be $4500 \times 4500$ pixels. If stored in uncompressed 16-bit per pixel (bpp) format, it would take about 40 MB for storage, and would take approximately half an hour for transmission using a high-speed modem [ZBI06]. Thus compression will play an increasingly important role in Picture Archiving and Communication Systems (PACS) to reduce file sizes while maintaining important diagnostic information.

In recent years, there has been discussion about which type of compression techniques, lossy or lossless, is better for mammogram compression. Although current lossless compression methods provide very high-quality images, the compression ratios are very low, typically from 1.5:1 to 4:1 [SKK+02, GAJI09]. On the other hand, several lossy compression methods provide acceptable compression ratios but come with considerable loss of image quality, which impacts diagnostic information [GSSMG00, LSM06, MKG08, PCG+97, PPM+97, Sch06, ZBI06].

Region-based compression techniques provide a good compromise between lossy and lossless compression [Koc83, GSSMG00, Gri01, CSSGG04]. Region-based compression combines lossless and lossy compression together by segmentation. After segmenting the images into the region-of-interest (ROI) and background, region-based compression compresses the ROI loss-
lessly to preserve important information, while compresses the background lossily to obtain a high compression ratio. Grinstead et al. [GSSMG00] introduce a quadtree fractal encoding scheme to generate Focus of Attention Regions (FARs), and then use a combination of lossy and lossless compression to provide a compression ratio of about 14:1 on mammograms. Chan et al. [CSSGG04] propose a content-based compression method using the fractal-based segmentation together with a modified JPEG2000. Their method can achieve an overall compression up to 30:1 (2:1 for the lossless compression of the ROI and 80:1 for the lossy compression) while fully preserving over 90% of the marked micro-calcification on mammograms.

In this thesis I describe a novel hybrid compression method for mammograms using super-resolution. I first use automatic region-growing segmentation to identify the clinically important areas. Then, a modified version of Huffman coding-based JPEG lossless compression is applied in such a way that the extracted areas from the first step are compressed losslessly, while the remaining regions are compressed lossily by downsampling before applying the encoding procedure, and applying Super-Resolution (SR) techniques after the decoding procedure to recover the original resolution image. The hybrid compression scheme can achieve an overall compression ratio of about 62:1 while fully preserving about 90% of the marked micro-calcification. Additionally, the micro-calcification detection system can obtain almost the same detection rates on our decompressed mammograms as on original mammograms.

6.2 Automatic segmentation

According to radiologists, the region of interest includes masses, microcalcifications, ducts, and the breast boundary, usually contains around 15% of the mammogram [GSSMG00]. Thus a successful segmentation of the regions of interest (ROIs) from the background and uninteresting area will help us develop a compression method that is a combination of lossy (for background and uninteresting regions) and lossless (for ROIs) to achieve high compression ratio while maintaining the important information for diagnosis.

For hybrid compression of mammograms, prior to compression, the mammogram is auto-
matically segmented into three non-overlapping regions: (1) background region that is basically black area, (2) the uninteresting area that contains little diagnostic information, and (3) the ROIs that contains diagnostic information. (Figure 6.1)

Then, a modified version of Huffman coding-based JPEG lossless compression is applied in such a way that the extracted ROIs are compressed losslessly, while the remaining regions are compressed lossily with super-resolution.

![Figure 6.1: Automatic segmentation and compression of mammograms](image)

### 6.2.1 Global threshold segmentation

Based on global information, the image histogram, I segment mammograms into two regions: background and breast region (Figure 6.2). The breast region usually has greater intensity than the background. Through experiments, I select a global threshold value to include as much the breast boundary pixels as possible. After finding the global threshold value I can then segment breast region against background. However global thresholding is not good for further segmentation of breast region because masses, microcalifications, blood vessels, and ducts usually have the same
intensity. Furthermore, some mammograms may contain labels which have similar intensity as breasts. I observe that the label regions are small compared with breast region. Therefore I use erosion and dilation to remove label regions on the background.

Figure 6.2: Global Threshold segmentation: (a) Original mammogram (b) Segmented mammogram after applying global threshold. (c) Segmented mammogram after removing label regions.

6.2.2 Region-growing segmentation

To further segment the breast region into region of interests and uninteresting area, I use a region-growing method (Figure 6.3).

The mammogram to be segmented differs greatly in contrast from the background image. Changes in contrast can be detected by operators that calculate the gradient of an image. The gradient image can be calculated and a threshold can be applied to create a binary mask containing the segmented breast. In this chapter, I use Sobel operator to obtain a binary mask that contains the seed pixels for region-growing.

Then I dilate the seed pixels using linear structuring elements. The dilated gradient mask shows the outline of the breasts quite nicely, but there are still holes in the interior of the breast. To fill these holes I first perform a flood-fill operation on background pixels of the binary im-
age. A hole is then found as a set of background pixels that cannot be reached by filling in the background. And then we flood-fill the holes.

![Figure 6.3: Region-growing Segmentation: (a) Input mammogram. (b) Binary gradient mask. (c) Dilated mask. (d) Filled mask. (e) Eroded mask. (f) Region of interest.](image)

### 6.3 Huffman coding-based JPEG near lossless compression

The general procedure of image compression can be simplified as a three-step procedure. At the first step, the intensities of the original images are transformed with a de-correlating operation, such as wavelets and Discrete Cosine Transform (DCT). Then the transform coefficients are quantized. At the last step, the quantized coefficients are coded with a lossless encoder. Mathematically, this procedure can be described as follows

\[
c = H[Q[Tb]]
\]

where \( b \) is a \((RC) \times 1\) vector that transformed from an input \( R \times C \) pixel image, \( T \) is an \((RC) \times (RC)\) transformation matrix, \( Q[.] \) is a quantization operator, \( H \) a lossless encoder that encode the quantized coefficients into entropy coded stream, and \( c \) is an \((RC) \times 1\) vector that contains the entropy coded stream. For decoding, the quantized coefficients are extracted from encoding, then the estimate of the original image is generated as follows
\[ \hat{b} = T^{-1} Q^*[H^*[c]] \]

where \( \hat{b} \) is the estimate of the original image, \( T^{-1} \) is the inverse of the transform operator, and \( Q^*[\cdot] \) represent a de-quantization operator, and \( H^*[\cdot] \) indicates decoding. Since coding is lossless, we have \( H^*[H[x]] = x \). However, quantization is a usually lossy procedure, so that \( Q^*[Q[x]] \neq x \).

### 6.3.1 JPEG compression

The name JPEG stands for Joint Photographic Experts Group, the name of the committee that created the JPEG standard and also other standards. The encoding process of JPEG consists of several steps [Jeo97]:

1. Convert input image from \( RGB \) color space to \( YC_bC_r \) color space, consisting of one luma component \( (Y) \), representing brightness, and two chroma components \( (C_b \text{ and } C_r) \), representing color.

2. Reduce the resolution of the chroma data by a factor of 2. Because human beings are less sensitive to color details than to brightness details. For monochrome images this step is usually skipped.

3. Split the image into \( 8 \times 8 \) pixel block, and for each block, apply discrete cosine transform (DCT) on each of the \( Y, C_b, \text{ and } C_r \) data.

4. Quantize \( DCT \) transform coefficients. The magnitudes of the high-frequency components are stored with a lower accuracy than the low-frequency components. A typical quantization matrix \( Q_{JP} \), as specified in the original JPEG Standard, is as follows:
5. Encode the resulting data for all $8 \times 8$ blocks with a lossless Huffman encoding.

The decoding process just reverses these steps. Mathematically, the JPEG encoding and decoding process can be described as follows

$$c_n = H_{Huf}[Q_{JPG}[T_{DCT}b_n]]$$

$$\hat{b}_n = T_{DCT}^{-1}Q_{JPG}^*[H_{Huf}^*[c_n]]$$

where $\hat{b}_n$ is the estimate of the original $8 \times 8$ block $b_n, c_n$ is the Huffman encoded stream, $T_{DCT}^{-1}$ is the inverse of the DCT operator $T_{DCT}, Q_{JPG}[.]$ is a quantization operator using a typical quantization matrix $Q_{JPG}$ as specified in the original JPEG Standard, $Q_{JPG}^*[.]$ represents the corresponding de-quantization operator, and $H_{Huf}[.]$ and $H_{Huf}^*[.]$ indicate Huffman encoding and decoding.

### 6.3.2 Huffman coding-based JPEG near lossless compression

For the region of interests, I use a Huffman coding-based JPEG near lossless compression. The compression process is very similar to JPEG. But instead of using the quantization matrix $Q_{JPG}$, I specify the minimum number of DCT coefficients used in an $8 \times 8$ block for Huffman encoding.
that restricts the compression artifacts in terms of Root Mean Square Error (RMSE) within a user specified amount \(\epsilon_t\). The procedure can be expressed as

\[
\hat{b}_n^k = T_{DCT}^{-1}[H_{Huf}^*[H_{Huf}[P_k[T_{DCT}b_n]]]]
\]

\[
\hat{k} = \arg \min_k \{RMSE(\sum_{n=1}^{N}\hat{b}_n^k, \sum_{n=1}^{N}b_n) \leq \epsilon_t\}
\]

\[
RMSE(X,Y) = \sqrt{MSE(X,Y)}
\]
\[
= \sqrt{E((X - Y)^2)}
\]
\[
= \sqrt{\frac{\sum_{i=1}^{T}(x_i - y_i)^2}{T}}
\]

where \(\hat{k}\) is the specified minimum number of DCT coefficients used in an \(8 \times 8\) block for Huffman encoding, \(b_n\) is the original \(8 \times 8\) block, \(\hat{b}_n^k\) is the estimate of the original \(8 \times 8\) block \(b_n\) after compression with \(k\) DCT coefficients, \(N\) is the total number of blocks in the image, \(T_{DCT}^{-1}\) is the inverse of the DCT operator \(T_{DCT}\), \(H_{Huf}[.\] and \(H_{Huf}^*[.\] indicate Huffman encoding and decoding, \(P_k[.\] is a selection function that selects the first \(k\) DCT coefficients used in an \(8 \times 8\) block for Huffman coding, and \(RMSE(X,Y)\) is the root mean square error between the two images \(X\) and \(Y\), and \(T\) is the total number of pixels in each image.

6.4 Compression using super-resolution

Incorporating super-resolution into compression aims to maximize compression ratio while maintaining relatively high-quality images. The flowchart of the proposed method is shown in Figure 6.4. The method first downsamples a high-resolution image \(HR(x)\) to a low-resolution one \(LR(x)\), then it uses an algorithm similar to JPEG to encode it. The decompression process first decodes the stored low resolution image to obtain \(LR'(x)\) and then applies a super-resolution algorithm to produce \(HR'(x)\).
In the pre-processing stage, we downsample the mammograms using bilinear downsampling. In the post-processing stage, the super-resolution algorithm generates high-resolution mammograms from low-resolution decoded mammograms with no manual registration. The super-resolution algorithm consists of four main steps. The first step automatically aligns the breasts to a standardized position. The second step uses a process called eigentransformation to infer a global model representing the low-frequency information in the image. In eigentransformation, Principal Component Analysis (PCA) is used to fit the input images as a linear combination of the low resolution images in the training set. The HR images are then inferred by replacing the LR training images with HR ones, while retaining the same combination coefficients. In the third step, a patch-based one-pass algorithm generates the high-frequency contents of the HR images. The fourth step remaps the breasts back to their original position. The details of eigentransformation and one-pass algorithm are described in chapter 3.
6.5 Evaluation

In this chapter, the quality of compressed image is defined as the similarity of the decompressed image with the original high-resolution image. We use Peak Signal-to-Noise Ratio (\(PSNR\)) and Mean Structural Similarity (\(MSSIM\)) index to measure the quality of results.

Let \(X\) and \(Y\) be two images to be compared. \(N\) is the total number of pixels in each image. The \(PSNR\), which is most commonly used as a measure of quality of reconstruction, is defined as

\[
PSNR(X, Y) = 20 \times \log_{10} \frac{255}{RMSE(X, Y)}
\]

where \(RMSE\) is the root mean square error between the two images.

\[
RMSE(X, Y) = \sqrt{MSE(X, Y)}
= \sqrt{E((X - Y)^2)}
= \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}}
\]

The structural similarity (SSIM) index [WBSS04] is an implementation of the idea of structural similarity, from an image formation point of view, which takes into account contrast, luminance, and structure to determine similarity between two images. \(SSIM\) is defined as

\[
SSIM(x, y) = \frac{(2\mu_x\mu_y + C1)(2\sigma_{xy} + C2)}{\mu_x^2 + \mu_y^2 + C1(\sigma_x^2 + \sigma_y^2 + C2)},
\]

\[
\sigma_{xy} = \frac{1}{T-1} \sum_{i=1}^{T} (x_i - \mu_x)(y_i - \mu_y)
\]

where \(x\) and \(y\) are subimages of \(X\) and \(Y\), \(T\) is the total number of pixels in each subimage, \(\mu_x\) is the average of \(x\), \(\mu_y\) is the average of \(y\), \(\sigma_x\) is the standard deviation of \(x\), \(\sigma_y\) is the standard deviation of \(y\). \(C1 = (k_1L)^2\) and \(C2 = (k_2L)^2\) are two variables to stabilize the division with small denominators, \(L\) is the dynamic range of the pixel values (typically this is 255), \(k_1 = 0.01\)
and $k_2 = 0.03$ by default. The mean SSIM (MSSIM) is then simply the mean of the SSIMs for all subimages. A value of MSSIM of 1 indicates perfect similarity [WBSS04].

The average compression ratio (ACR) of hybrid compression is computed as follows:

$$ACR = \left( \frac{R_1}{BCR} + \frac{R_2}{LCR} + \frac{R_3}{LLCR} \right)^{-1}$$

where $R_1$ is the percentage of the image contained in the background area, $R_2$ is the percentage of the image contained in the uninteresting area, and $R_3$ is the percentage of the image contained in the ROI, $BCR$ is the compression ratio of the background, $LCR$ is the lossy compression ratio of uninteresting area, and $LLCR$ is the lossless compression ratio of the ROI.

Furthermore, super-resolution uses statistical machine learning techniques to learn the function that maps low resolution images to their high resolution counterparts, thus some details of the decompressed images may be different from the original images. To determine whether hybrid compression affects clinical diagnostic performance, I use texture-based features with neural networks to automatically detect micro-calcification on both original and decompressed digital mammograms. The first step is to enhance the micro-calcification with non-subsampled contourlet transform as in [CZD06], since micro-calcification are usually very small in size. Then we use Gray-Level Difference Method (GLDM) [KP99] to extract the texture-based features for detection. GLDM is based on the occurrence of two pixels which have a given absolute difference in pixel value and which are separated by a specific displacement $\theta$. The definition of GLDM is as follows.

$$\theta = (\Delta x, \Delta y)$$

$$D_\theta(x, y) = |D(x, y) - D(x + \Delta x, y + \Delta y)|$$

$$P(i|\theta) = \text{Prob}(D_\theta(x, y) = i), 0 \leq i \leq 255$$

$$CDF(t) = \sum_{i=0}^{t} P(i|\theta)$$

where $\theta$ is the given displacement vector, and the four possible forms of the displacement vector will be considered are $(0, s), (-s, s), (s, 0)$, and $(-s, -s)$ where $s$ is the inter-sample
spacing, $D_0(x, y)$ is the absolute difference in pixel value of two given pixels, $P(i|\theta)$ is the estimated Probability-Density Function (PDF), $i$ is a pixel value between 0 and 255, $CDF(t)$ is the Cumulative Distribution Function (CDF) for a given pixel value $t$.

Finally we use a four-layer back-propagation neural network as the classifier. The CDF from GLDM is used as input signal of the input layer. The output layer just contains one single node for classification into positive or negative state. I use a nonlinear sigmoid function as transfer function for each neuron, and use scaled conjugate gradient method for training [Mol93].

Figure 6.5 shows some sample mammograms with labels of micro-calcification from the MIAS dataset, the window that contains the micro-calcification, and the corresponding window enhanced by Nonsubsampled Contourlet Transform. The location of the micro-calcification in mammograms of MIAS is given by radiologists.

Figure 6.6 shows the texture feature of Gray-Level Difference Method (GLDM) and its four Cumulative Distribution Functions (CDFs) for four different displacement vectors of the input image.

### 6.6 Experimental results

I use DDSM (Digital Database for Screening Mammography) [HBK+98][HBK+01] and mini-MIAS dataset of mammograms [SPD+94] for experiments. DDSM is a standard dataset used by the mammography image analysis research community. The database has about 2,500 cases. Each case includes two images of each breast, along with some associated patient information and image information. In this chapter, randomly selected 400 normal left mediolateral oblique (MLO) images from the DDSM dataset are used for training and randomly selected 10 normal left MLO images for testing. We use an automatically built $32 \times 8$ mesh to register the mammograms.

MIAS dataset has about 161 cases. Each case includes a pair of images that represent the left and right mammograms of a patient, along with some associated patient and image information. The size of all the images is $1024 \times 1024$ pixels. The images have been centered in the matrix with edge padded. In this chapter, I use randomly selected 10 left MLO images from the MIAS dataset
for training and testing. Though in this chapter, I tested only on left MLO views. This method can be applied to right MLO views and Craniocaudal (CC) views with little modifications.

Figures 6.7 and 6.8 show the image quality measures for compression with super-resolution algorithm on DDSM. The downsample factors in this experiment are 4, 8, 16, and the number of Discrete Cosine Transform (DCT) coefficients used in an $8 \times 8$ block for JPEG encoding are 1, 2, 3, 5, 10, and 15. Table 6.1 compares the results of compression with super-resolution (downsample factor = 16, 15 DCT coefficients) with the results of JPEG2000 and lossless JPEG. Sample results of compression with super-resolution of different resolutions are reported in Figure 6.10.

Figure 6.9 shows a subtraction of an original image from the corresponding decompressed image after hybrid compression to give more intuitive description of the differences in structure.
Figure 6.6: Gray-Level Difference Method (GLDM): Four Cumulative Distribution Functions (CDFs) for four different displacement vectors of the input image.

From the results we can see that main differences between the original image and the corresponding decompressed image are in the uninterested area.

<table>
<thead>
<tr>
<th>Compression methods</th>
<th>PSNR</th>
<th>Compression ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPEG2000</td>
<td>41.95</td>
<td>80:1</td>
</tr>
<tr>
<td>Lossless JPEG</td>
<td>lossless</td>
<td>3:1</td>
</tr>
<tr>
<td>Compression with SR</td>
<td>35.9023</td>
<td>1792:1</td>
</tr>
</tbody>
</table>

Table 6.1: Average results of compression with super-resolution for 10 mammograms compared with JPEG 2000 and lossless JPEG

From figures 6.7 and 6.8, we can see that when we just use 1 or 2 DCT coefficients, which results in very blocky images, we get a PSNR of 33.458 after super-resolution of this kind of blocky images. These results indicate that super-resolution can also attenuate the JPEG compression artifacts when the compression ratio is high. From Table 6.1, we can see that compared
Figure 6.7: PSNR of the decompressed images after super-resolution with the results of JPEG2000 and lossless JPEG, the compression ratio of compression using super-resolution method is hundreds of times higher than lossless JPEG and JPEG 2000 with a slightly lower PSNR.

Although the PSNR is a physically meaningful metric for signal reconstruction, it does not necessarily reflect perceived visual quality by humans [Gir93]. From figure 6.10, we can observe image artifacts for mammograms compressed with downsampling factor 16 with 10 and 15 DCT coefficients. However, from figures 6.7 the corresponding PSNR is equal or better than those compressed with downsampling factor 4. Therefore, I also use the mean Structural Similarity (SSIM) Index, which aims to primarily measure the structural changes between a reference image and its distorted version [WBSS04].

I also test the Huffman coding-based JPEG lossless compression on MIAS dataset. The number of DCT coefficients tested in an $8 \times 8$ block for Huffman encoding are 1, 2, 4, 8, 16, 32, and 64. We report the corresponding compression ratio and PSNR in Table 6.3. Sample results
Finally, the results of hybrid compression on MIAS dataset is reported in table 6.4. To find the regions of interest, I use region-growing segmentation. The downsample factor for super-resolution is 16, and we use 15 DCT coefficients for Huffman coding-based JPEG near lossless compression. The average compression ratio of hybrid compression is $62 : 1$ which is 20 times higher than lossless JPEG while having the same PSNR in the region of interests. Sample results are reported in Figure 6.13 (c). A close observation of a local area of Figure 6.13 is illustrated in Figure 6.14.

The accuracy of automatic segmentation is very important for hybrid mammogram compression, since both the compression accuracy and compression ratio will largely depend on the segmentation. This thesis tested the accuracy of region-growing segmentation by computing the micro-calcification coverage rates on MIAS dataset. I use 330 mammograms in total which contain 26 cases of micro-calcifications. Figure 6.11 shows the sample result of auto-

![Image](image.png)

Figure 6.8: MSSIM of the decompressed images after super-resolution
matic region-growing segmentation. The red area is the ROI which includes important clinical information, and the blue circle encloses micro-calcification, according to the information given by MIAS dataset. Figure 6.12 gives the micro-calcification coverage results of the automatic region-growing segmented mammograms, in terms of recall and F-measure.

To determine whether hybrid compression artifacts would affect clinical diagnostic performance, I apply an automatic micro-calcification detection system on both original and decompressed mammograms. This thesis uses about 5000 windows with micro-calcification and 5000 windows without micro-calcification for training. For testing, I use 10-fold cross validation. Since the MIAS dataset only contains about 30 mammograms that contain micro-calcification, to balance the training dataset, I generate the rest of the positive training samples by rotating, scaling, and mirroring of the existing samples. The negative training samples are generated by scanning a window across mammograms without micro-calcification. Table 6.2 gives the micro-calcification detection results on both original mammograms and decompressed mammograms. From the results we can see that we can get almost the same detection rates on decompressed mammograms as on original mammograms.
<table>
<thead>
<tr>
<th>Measures</th>
<th>Original</th>
<th>Decompressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.9881</td>
<td>0.9839</td>
</tr>
<tr>
<td>Recall</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>F-measure</td>
<td>0.994</td>
<td>0.9919</td>
</tr>
</tbody>
</table>

Table 6.2: 10-fold cross validation micro-calcification detection results on MIAS dataset

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Compression ratio</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>5.7111</td>
<td>303.0913:1</td>
</tr>
<tr>
<td>32</td>
<td>6.1752</td>
<td>52.2291:1</td>
</tr>
<tr>
<td>16</td>
<td>6.6955</td>
<td>46.2037:1</td>
</tr>
<tr>
<td>8</td>
<td>7.1026</td>
<td>39.0931:1</td>
</tr>
<tr>
<td>4</td>
<td>7.4043</td>
<td>35.2940:1</td>
</tr>
<tr>
<td>2</td>
<td>7.6626</td>
<td>32.3263:1</td>
</tr>
<tr>
<td>1</td>
<td>7.7841</td>
<td>29.5642:1</td>
</tr>
</tbody>
</table>

Table 6.3: Huffman coding-based JPEG near lossless compression results on MIAS dataset

6.7 Chapter summary

In this chapter I describe a novel hybrid compression method for mammograms. I first use automatic segmentation to identify the clinically important areas. Then, a modified version of Huffman coding-based JPEG lossless compression is applied in such a way that the extracted areas from the first step are compressed losslessly, while the remaining regions are compressed lossily by downsampling before applying the encoding procedure, and applying super-resolution techniques after the decoding procedure to recover the original resolution image.

Though the compressed images have very high-resolution, some details in the compressed images are different from the original images in the uninteresting region. To further determine whether compression affects clinical diagnostic performance, I will study whether these differ-
<table>
<thead>
<tr>
<th>Region</th>
<th>Area</th>
<th>Compression ratio</th>
<th>Method</th>
<th>PSNR</th>
<th>MSSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65.65%</td>
<td>5507100:1</td>
<td>1 bit</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>2</td>
<td>24.7%</td>
<td>1792:1</td>
<td>Compression with SR</td>
<td>24.29</td>
<td>0.9167</td>
</tr>
<tr>
<td>3</td>
<td>9.65%</td>
<td>6:1</td>
<td>Near Lossless JPEG</td>
<td>303.0193</td>
<td>1</td>
</tr>
<tr>
<td>Average compression ratio</td>
<td>62:1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.4: Hybrid compression results on MIAS dataset

ences would affect breast cancer diagnosis by both radiologists and applying an computer-aided breast cancer detection system to the original images and compressed images respectively, and then compare the detection rates.
Figure 6.10: Sample results of compression with super-resolution of different resolutions. (a) Original image. (b) The decompressed images after super-resolution (Coeffs=1,2,3,5,10,15; downsample factor=4). (c) The decompressed images after super-resolution (Coeffs=1,2,3,5,10,15; downsample factor=8). (d) The decompressed images after super-resolution (Coeffs=1,2,3,5,10,15; downsample factor=16).
Figure 6.11: Sample result of region-growing segmentation. The red area is the ROI which includes important clinical information, and the blue circle encloses micro-calcification
Figure 6.12: Micro-calcification coverage results of automatic region-growing segmentation.
Figure 6.13: Sample results of hybrid compression. (a) Original image. (b) JPEG compressed image (The number of DCT coefficients used for compression from top to bottom is 1, 2, 4, 8, 16, 32, and 64 respectively). (c) Compressed image using hybrid compression (Coeffs=15, downsampling factor=16).
Figure 6.14: A close observation of the sample results of hybrid compression. (a) Original image. (b) JPEG compressed image (The number of DCT coefficients used for compression from top to bottom is 1, 2, 4, 8, 16, 32, and 64 respectively). (c) Compressed image using hybrid compression (Coeffs=15, downsample factor=16).
Chapter 7

Specular Normal Synthesis Using
Stochastic Super-resolution for Detailed
Facial Geometry

7.1 Introduction

Detailed facial geometry contributes significantly to the visual realism of face models in computer
games, movies, virtual reality applications, and so on.

In this chapter I describe a new technique for acquiring specular normal maps for detailed
facial geometry using spherical gradient illumination as in [ZG10]. The key elements of the
approach are the use of stochastic super-resolution to generate specular normal map based on
diffuse normal map, instead of capturing both of them. I analyze a training dataset of diffuse nor-
mal maps and specular normals of a particular object and learn the mapping from low-frequency
components of diffuse normal maps to high-frequency components of specular normal maps of
that object. This enables us to infer the specular normal map detail depicting the same person as
a diffuse normal map given as input. Experimental results show that the proposed algorithm gen-
erates high-quality specular normal maps from diffuse normal map inputs, and demonstrate the
synthesized normal map can be used for image-based lighting. I also demonstrate the effective-
ness of this method in several applications, including transferring specular normal information
from one person to other people, and transferring specular normal information to unpolarized
data.
7.2 Method

The facial scanning system captures the face geometry by combining a medium-resolution structured light scan base geometry with high-resolution surface normals. The structured-light scanning generates the base geometry. The surface normals are obtained from gradient illumination and polarization. Finally the mesh vertices of the base geometry are optimized to match the normal maps using an embossing technique.

7.2.1 Hardware setup

The structured-light-scanning system uses a stereo pair of SLR cameras plus a projector placed between them (Figure 7.1). The projector projects a sequence of four colored strip patterns and one uniform white pattern, with a long throw lens to concentrate the available light (Figure 7.2).

Figure 7.1: Setup hardware for face scanning.

The primary light apparatus to create the gradient lighting patterns consists of about 150 LED lights placed on the vertices and edges of a twice-divided icosahedron (Figure 7.3). For polarized patterns, individual linear polarizers are placed over each LED light. A linear polarizers is mounted on a servomotor in front the camera, which enable to polarizer to be rapidly flipped
Figure 7.2: The projector projects a sequence of four colored strip patterns and one uniform white pattern.

on its diagonal between horizontal and vertical orientations (Figure 7.9). A vertical polarizer is placed in front of the LCD projector as described in [MHP+07].

In capturing, I take 8 photos for each gradient pattern under two linear polarization states (Figure 7.8), and 5 pairs of stereo photos for each structures light strip patterns (Figure 7.6), using Cannon 5D cameras in burst mode which requires just a few seconds to capture data at 12 megapixel resolution [MHP+07].

7.2.2 Camera calibration

The intrinsic and extrinsic parameters of the cameras are calibrated using a technique as in [Zha00]. Camera calibration is very important in 3D face scanning. The calibration process uses the camera to observe a planar shown at a few different orientations (Figure 7.4), then detects the feature points in the captured images, and uses the closed-form solution as described in [Zha00], followed by a nonlinear refinement based on the maximum likelihood criterion. The planar can be moved freely (Figure 7.5). Compared with conventional camera calibration techniques which require expensive equipment such as two or three orthogonal planes, the technique
Figure 7.3: The primary light apparatus we use to create the gradient lighting patterns consists of about 150 LED lights placed on the vertices and edges of a twice-divided icosahedron.

is simple and robust.

7.2.3 Geometry processing

Because of noise and the limited resolution of the projector, the structured light scan introduces some high frequencies biasing and noise. I smooth the structured light scan surface using bilateral denoising, and then create a surface normal map from the smoothed mesh and extract the high frequency details of the estimated normals using high-pass filtering [MHP+07].

The most important part of stereo algorithm is matching correspondences in two stereo images (Figure 7.6). The correspondences are computed between the cameras by finding best matches of structure light pattern images $C_t$ ($t = 1$ to $N_{\text{patterns}}$). I use the ratio images $R_t$ for stereo matching which have the pure intensity ratio from the projector.

$$R_t = \frac{S_t}{U}, (t = 1, ..., N_{\text{patterns}})$$
Figure 7.4: (a) The camera calibration process uses the camera to observe a planar shown at a few different orientations. (b) The planar for calibration

where $R_k$ is the ratio image for colored strip pattern $t$, $C_t$ is the stereo image for colored strip pattern $t$, and $U$ is the stereo image for the uniform white pattern. Figure 7.7 illustrates the ratio images of the stereo pair. Then a dynamic programming method as described in [ZCS02] for stereo matching is implemented. And, by matching the pixels between the sensor images, we can reconstruct the surface through triangulation.

7.2.4 Normal map processing

A normal map is usually an RGB image that corresponds to the X, Y, and Z coordinates of a surface normal. To compute normal maps from gradient patterns, I use the algorithm as described in [MHP+07, GCP+09] to process the gradient illumination images whose reflectance is either diffuse or specular. Figure 7.8 illustrates the 8 photos for each gradient pattern under two linear polarization states.

Because most surfaces exhibit a combination of specular and diffuse reflectance, we can estimate the specular and diffuse normals independently. As the polarization state of the specular
Figure 7.5: Move the planar freely to show at a few different orientations.

Reflectance is determined by the polarization state of the incident light, while diffuse reflectance arises from subsurface scattering and is almost completely unpolarized regardless of the polarization characteristics of the incident light, we can separate the diffuse and specular reflectance by controlling the polarization state of the incident light and measuring the polarization state of the reflected light [MHP\(^+\)07, GCP\(^+\)09]. To separate diffuse and specular reflectance I place a vertical polarizer over the light source and a linear polarizer is mounted on a servomotor in front the camera, which enables the polarizer to be rapidly flipped on its diagonal between horizontal and vertical orientations (Figure 7.9). When the polarizer is in its horizontal orientation, all of the specularly reflected light, and half of the diffusely reflected light are blocked from the camera, producing an image \(I_1 = \frac{1}{2}I_D\). When the polarizer is in its vertical orientation, the camera produces \(I_2 = \frac{1}{2}I_D + I_S\). The diffuse and specular image components are then \(I_D = 2I_1\) and \(I_S = I_2 - I_1\) [MHP\(^+\)07, GCP\(^+\)09]. The diffuse-specular separation using linear polarization is illustrated in Figure 7.8.
To compute the normals from the 4 gradient illumination images, I divide the gradient images by the constant image, and then map the range to a normal direction. Figure 7.10 shows the processed specular and diffuse normal map.

### 7.2.5 Geometry optimization

At this point, we have acquired a base geometry through structured-light scanning, and diffuse and specular normal maps from gradient illumination and polarization. However, the base geometry is inaccurate due to translucency of human skin. Then I optimize the base mesh vertices to match the normal maps using an embossing process as in [NRDR05]. I first smooth the base geometry using bilateral denoising as in [MHP+07], and create a surface normal map from the smoothed mesh. Then I extract the high frequency details of the diffuse and specular normal maps using high-pass filtering and add these details to the smoothed geometric normals to obtain
Figure 7.7: Structured light ratio images. (Top) Images captured by the left camera. (Bottom) Images captured by the right camera.

a hybrid normal map. Lastly, I optimize the mesh vertices using a greedy algorithm to match the hybrid normal maps as in [NRDR05]. Generally, the geometry optimization consists of 5 steps:

1. Smooth the base geometry using bilateral denoising.
2. Create a surface normal map $N_s$ from the smoothed mesh.
3. Extract the high frequency details $H_d$ or $H_s$ of the diffuse or specular normal maps using high-pass filtering.
4. Compute the hybrid normal map $N_a$ by adding the details $H_d$ or $H_s$ to the smoothed geometric normals $N_s$. $N_a = N_s + H_d$ or $N_a = N_s + H_s$.
5. Optimize the mesh vertices using a greedy algorithm to match $N_a$.
7.2.6 Specular normal map synthesis

In this thesis, I go further than previous work by Ma [MHP+07] for acquiring specular normal map by synthesizing the specular normals from diffuse normals instead of capturing both of them.

Given a diffuse normal map as input, to construct the corresponding specular normal map, we first filter the diffuse normal map with a normal map filter similar to Gaussian high-pass filter [HSRG07], and then subdivide the filtered diffuse normal map into patches, which we call the low-frequency patches, by scanning a window across the image in raster-scan order. Similarly, I also filter and subdivide the specular normal map in the training set into patches which we call high-frequency patches.

To synthesize a specular normal map, for each low-frequency patch, a high-frequency patch of the training specular normal map is selected by a stochastic search from the training set based on local diffuse normal details and adjacent, previously determined specular normal map patches. The selected high-frequency patch should not only come from a location in the training images that has a similar corresponding low-frequency appearance, but it should also match at the edges of the high-frequency patch with the overlapping pixels, which we call high-frequency overlap, of its previously determined high-frequency neighbors to ensure that the high-frequency patches are compatible with those of the neighboring high-frequency patches.

In this chapter I synthesize the specular normal map with an algorithm that is an extension of the one-pass algorithm, proposed by Freeman et al. [FJP02, FPC00]. In the one pass algorithm, I first concatenate the pixels in the low-frequency patch and the high-frequency overlap to form a search vector. The training set also contains a set of such vectors. Then I search for a match by finding the best match in the training set. When I find a match we extract the corresponding high-frequency patch from training data set and add it to the initial diffuse normal map to obtain the output specular normal map. The details of one-pass algorithm are described in chapter 3.
7.3 Experimental results

For experiments I use a high dynamic range (HDR) face dataset from Institute for Creative Technologies of University of Southern California (Figure 7.11), which consists of floating-point images of $1944 \times 1296$ pixels with a frontal view of about 15 different expressions under 4 gradient lighting patterns and 5 structured-light patterns. The primary light apparatus we use to create the gradient lighting patterns consists of about 150 LED lights placed on the vertices and edges of a twice-divided icosahedron (Figure 7.3). For polarized patterns, individual linear polarizers are placed over each light. A linear polarizers is also mounted on a servomotor in front the camera, which enable the polarizer to be rapidly flipped on its diagonal between horizontal and vertical orientations to capture both diffuse and specular reflectance.

In preprocessing I register the normal maps using manually drawn facial masks (Figure 7.12), so that I can assume that the same parts of faces appear in roughly the same parts of the images. The diffuse and specular normal map size is fixed to $1944 \times 1296$ pixels and we use them as the training images.

To construct our training dataset, I filter the diffuse normal maps with a normal map filter similar to Gaussian high-pass filter. Then I subdivide the filtered images into low-frequency patches by scanning an $8 \times 8$ pixel window across the image in raster-scan order. Then I again filter and subdivide the training specular normal maps into $8 \times 8$ pixel high-frequency patches. At each step we also get a 17-pixel overlap of each high-frequency patch with the high-frequency patches above and to the left. Then I create the training vectors by concatenating the low-frequency patches and corresponding high-frequency overlaps. In practice, the size of low-frequency patches and high-frequency patches is not necessarily the same. The parameter $\alpha$, which controls the trade-off between matching the low-frequency patches and finding the most compatible high-frequency patches, is set to 0.2. The parameter $\beta$ balances the contribution of each color channel.

Figure 7.13 shows the specular normal map synthesis results of the same person with different expressions. From the results, we can see that the synthesized specular normal map has
much clearer detailed features than the diffuse normal map, and almost the same as the captured specular normal map.

Figure 7.14 shows the geometry optimization results of the same person with different expressions. In this figure, we present the geometry optimized with our synthesized specular normal maps. For comparison, we also present the base geometry captured through structured-light scan without any optimization, and the geometry optimized with the captured specular normal map.

Figure 7.15 shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map. The results show that our synthesized specular normal maps can also work with image-based lighting techniques.

Figure 7.16 to 7.18 show specular normal map synthesis results of cross-object transfer. The algorithm transfers specular information of our training dataset to another person.

Figure 7.19 to 7.21 show specular normal map synthesis results from unpolarized mixed normal map. The algorithm transfers specular information of our training dataset to unpolarized data. The results show that our algorithm can improve the quality of facial scanning without polarization by adding specular normal map information to unpolarized data.

7.4 Chapter summary

In this chapter I propose a new technique for acquiring specular normal map for high-resolution facial scanning using spherical gradient illumination. The key elements of the approach are the use of stochastic super-resolution to generate specular normal map based on diffuse normal map, instead of capturing both of them. I analyze a training dataset of diffuse normal maps and specular normal maps of a particular object and learn the mapping from low-frequency components of diffuse normal maps to high-frequency components of specular normal maps of that object. This enables us to infer, for example, the most likely high-frequency specular normal map detail depicting the same person as a low-resolution diffuse normal map given as input. Experimental results show that the proposed algorithm generates high-quality specular normal maps from diffuse normal map inputs.
Figure 7.8: Captured spherical gradient illumination images. (Top) Images with specular reflections. (Middle) Images with diffuse reflections. (Bottom) The four spherical gradient lighting conditions reflected in a mirrored sphere. (a) A linear gradient along the x-coordinate. (b) A linear gradient along the y-coordinate. (c) A linear gradient along the z-coordinate. (d) A constant pattern.
Figure 7.9: A linear polarizers is mounted on a servomotor in front the camera, which enables the polarizer to be rapidly flipped on its diagonal between horizontal and vertical orientations.

Figure 7.10: Normal acquisition from spherical gradient illumination. (a) Diffuse normal map. (b) Specular normal map.
Figure 7.11: A high dynamic range (HDR) face dataset which consists of floating-point images of about $1944 \times 1296$ pixels with a frontal view of about 15 different expressions.

Figure 7.12: Manually drawn facial masks. (a) Face image. (b) Facial mask.
Figure 7.13: (a) Diffuse normal map. (b) Synthesized specular normal map. (c) Captured specular normal map. This figure shows the specular normal map synthesis results of the same person with different expressions. The synthesized specular normal map has much clearer detailed features than diffuse normal map, and almost the same as the captured specular normal map.
Figure 7.14: (a) Base geometry. (b) Geometry optimized with synthesized specular normal map. (c) Geometry optimized with captured specular normal map. This figure shows the geometry optimization results of the same person with different expressions.
Figure 7.15: Renderings of face scans with different illumination sources. (a) Grace Cathedral. (b) Uffizi Gallery. (c) Pisa courtyard. This figure shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map. The results show that our synthesized specular normal maps can also work with image-based lighting techniques.
Figure 7.16: (a) Diffuse normal map. (b) Synthesized specular normal map. (c) Captured specular normal map. This figure shows the specular normal map synthesis results of cross-object transfer. The algorithm transfers one person’s specular information to another person.
Figure 7.17: (a) Base geometry. (b) Geometry optimized with synthesized specular normal map. (c) Geometry optimized with captured specular normal map. This figure shows the geometry optimization results of cross-object transfer.
Figure 7.18: Renderings of face scans with different illumination sources. (a) Grace Cathedral. (b) Uffizi Gallery. (c) Pisa courtyard. This figure shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map which is transferred from another person.
Figure 7.19: (a) Mixed normal map. (b) Synthesized specular normal map. (c) Captured specular normal map. This figure shows the specular normal map synthesis results of the same person with different expressions. The synthesized specular normal map are synthesized from unpolarized mixed normal map.
Figure 7.20: (a) Base geometry. (b) Geometry optimized with synthesized specular normal map. (c) Geometry optimized with captured mixed normal map. This figure shows the geometry optimization results of the same person with different expressions.
Figure 7.21: Renderings of face scans with different illumination sources. (a) Grace Cathedral. (b) Uffizi Gallery. (c) Pisa courtyard. This figure shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map. The results show that our synthesized specular normal maps can also work with image-based lighting techniques.
8.1 Conclusion

The driving motivation of the different applications addressed in this thesis is unified under the name of super-resolution. In this work I present algorithms that can enhance the resolution of degraded images according to specific requirements of the application for which the images are used.

One application addressed in this work is the problem of super-resolution of surveillance videos. In chapter 3, I improve the efficiency of face image super-resolution using stochastic search for local modeling. Experimental results show that the proposed algorithm generates high-quality face images from low-resolution inputs while reducing the computation time dramatically. Though face super-resolution could improve the appearance of face images dramatically, the detailed facial features such as eyes, eyebrows, nose, mouth and teeth of the super-resolution face are different from the ground truth. In chapter 4, I study whether face super-resolution can help face recognition by either computers or human beings, and propose a simultaneous face super-resolution and recognition algorithm. The experiments show that face super-resolution can not improve the performance of conventional face identification algorithms, but my method, which combines face super-resolution and face identification, can help face identification by computers in low-resolution images.

Another application addressed in this work is super-resolution of mammograms. I study mammogram super-resolution, which synthesizes a high-resolution mammogram from a low-resolution input automatically, with the help of a large collection of other high-resolution mammograms from many individuals, and study whether super-resolution can help automatic breast...
cancer detection. The experimental results show that the super-resolution algorithm can generate high-quality high-resolution breast mammographies from low-resolution input with no manual registration.

Additionally, this work addresses the problem of recovering a high-resolution image from an compressed image or a sequence of compressed images. I first downsample the image or image sequences as part of the pre-processing step before compression, and then apply super-resolution algorithms as post-processing of the compression process. The goal is to maximize the compression rate while maintaining the quality of the compressed images. The results show that the super-resolution for compression not only improves the spatial resolution of the degraded image but also reduces the compression artifacts.

Finally, this work proposes a new technique for real-time high-resolution 3D facial scanning using stochastic super-resolution to generate a specular normal map based on the diffuse normal map, instead of capturing both of them during scanning process. I analyze a training dataset of diffuse normal maps and specular normals of a particular object to learn the mapping from low-frequency components of diffuse normal maps to high-frequency components of specular normal maps of that object, and then infer the most likely high-resolution specular normal map detail depicting the same person as a low-resolution diffuse normal map given as input. Experimental results show that the proposed algorithm generates high-quality specular normal maps from diffuse normal map inputs.

8.2 Future work

The first area for future research is generic object super-resolution, for which a very large training dataset that contains a subset for each specific object is needed. It will also be interesting to exploit super-resolution for object recognition, both by computers and by people.

The second area of future research is to reduce the running time of the super-resolution algorithm. The drawback of this super-resolution algorithm is that it requires a very long time to perform the task. The time for training required by the algorithm is proportional to the size of
the training dataset. The system can not run in real-time. It will be interesting to use stream processing to parallelize the execution. Stream processing permits the execution of data-parallel algorithms with stream processors such as graphic processing units (GPUs), while using the central processing unit (CPU) for other purposes simultaneously. This would enable a conventional PC to run the image super-resolution algorithm in real time.

The third area of future research is automatic registration. The super-resolution algorithm requires the images to be well registered. Right now it is hard to do the registration automatically for face images.

Though the SR images have very high-resolution, some details in the SR images are different from the original images. The fourth area of future research is to further determine whether the differences would affect clinical diagnostic performance in medical imaging. It will be interesting study whether these differences would affect breast cancer diagnosis by both radiologists and applying an computer-aided breast cancer detection system to the original images and compressed images respectively, and then compare the detection rates.
References


Curriculum Vitae

Jun Zheng was born on October 14, 1981. The only son of Ganwu Mao and Yaxian Zheng. He graduated from The First Middle School of Changsha, Hunan, People’s Republic of China, in the spring of 2000. He entered Central South University, Hunan, People’s Republic of China, in the fall of 2000, and received his bachelor’s degree of Electronic Engineering in the spring of 2004, and later received his master’s degree of Electronic Engineering in the spring of 2007.

In the fall of 2007, he entered the Graduate School of The University of Texas at El Paso (UTEP). While pursuing a Ph.D. degree in Computer Science he worked as a Research Assistant for the Vision and Learning Lab (VL2). He was a member of the Association for Computing Machinery and UTEP chapter of Honor Society of Phi Kappa Phi.

Permanent address: 14257 Spanish Point Dr.

El Paso, Texas 79938